

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/53, 15/55, 15/60, C12P 7/18,</b> <b>C12N 9/04, 9/16, 9/88</b>	<b>A2</b>	<b>(11) International Publication Number: WO 98/21341</b> <b>(43) International Publication Date: 22 May 1998 (22.05.98)</b>
<b>(21) International Application Number:</b> PCT/US97/20873 <b>(22) International Filing Date:</b> 13 November 1997 (13.11.97)  <b>(30) Priority Data:</b> 60/030,601 13 November 1996 (13.11.96) US  <b>(71) Applicant (for all designated States except US):</b> GENENCOR INTERNATIONAL, INC. [US/US]; 4 Cambridge Place, 1870 South Winton Road, Rochester, NY 14618 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> DUNN-COLEMAN, Nigel, S. [GB/US]; 142 Johnson Avenue, Los Gatos, CA 95032 (US). DIAZ-TORRES, Maria [ES/US]; 58 North El Camino Real, San Mateo, CA 94401 (US). CHASE, Matthew, W. [US/US]; 2211-27 Hastings Drive, Belmont, CA 94002 (US). TRIMBUR, Donald [US/US]; 349 Orchard Avenue, Redwood City, CA 94601 (US).  <b>(74) Agent:</b> GLAISTER, Debra, J.; Genencor International, Inc., Page Mill Road, Palo Alto, CA 94304-1013 (US).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>Without international search report and to be republished upon receipt of that report.</i>
<b>(54) Title:</b> METHOD FOR THE RECOMBINANT PRODUCTION OF 1,3-PROPANEDIOL		
<b>(57) Abstract</b>  The present invention provides an improved method for the production of 1,3-propanediol from a variety of carbon sources is an organism comprising DNA encoding protein X of a dehydratase or protein X in combination with at least one of protein 1, protein 2 and protein 3. The protein X may be isolated from a diol dehydratase or a glycerol dehydratase. The present invention also provides host cells comprising protein X that are capable of increased production of 1,3-propanediol.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## METHOD FOR THE RECOMBINANT PRODUCTION OF 1,3-PROPANEDIOL

### Related Applications

5           The present application is a continuation-in-part application of United States Provisional Application 60/030,601 filed November 13, 1996, hereby incorporated herein in its entirety.

### Field of Invention

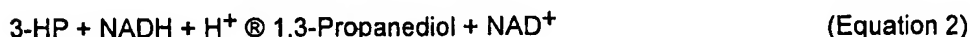
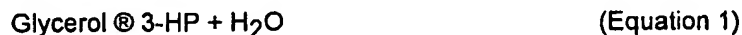
10           The present invention relates to the field of molecular biology and specifically to improved methods for the production of 1,3-propanediol in host cells. In particular, the present invention describes components of gene clusters associated with 1,3-propanediol production in host cells, including protein X, and protein 1, protein 2 and protein 3. More specifically the present invention describes the expression of cloned genes encoding protein X, protein 1, protein 2 and protein 3, either separately or together, for the enhanced production of 1,3-propanediol in host cells.

### Background

15           1,3-Propanediol is a monomer having potential utility in the production of polyester fibers and the manufacture of polyurethanes and cyclic compounds.

A variety of chemical routes to 1,3-propanediol are known. For example ethylene oxide may be converted to 1,3-propanediol over a catalyst in the presence of phosphine, water, carbon monoxide, hydrogen and an acid, by the catalytic solution phase hydration of acrolein followed by reduction, or from hydrocarbons such as glycerol, reacted in the presence of carbon monoxide and hydrogen over catalysts having atoms from group VIII of the periodic table. Although it is possible to generate 1,3-propanediol by these methods, they are expensive and generate waste streams containing environmental pollutants.

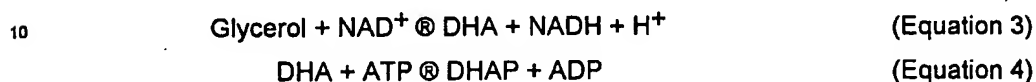
25           It has been known for over a century that 1,3-propanediol can be produced from the fermentation of glycerol. Bacterial strains able to produce 1,3-propanediol have been found, for example, in the groups *Citrobacter*, *Clostridium*, *Enterobacter*, *Ilyobacter*, *Klebsiella*, *Lactobacillus*, and *Pelobacter*. In each case studied, glycerol is converted to 1,3-propanediol in a two step, enzyme catalyzed reaction sequence. In the first step, a dehydratase catalyzes the conversion of glycerol to 3-hydroxypropionaldehyde (3-HP) and water (Equation 1). In the second step, 3-HP is reduced to 1,3-propanediol by a NAD<sup>+</sup>-linked oxidoreductase (Equation 2).



35           The 1,3-propanediol is not metabolized further and, as a result, accumulates in high concentration in the media. The overall reaction consumes a reducing equivalent in the form of a cofactor, reduced b-nicotinamide adenine dinucleotide (NADH), which is oxidized to nicotinamide adenine dinucleotide (NAD<sup>+</sup>).

-- 2 --

The production of 1,3-propanediol from glycerol is generally performed under anaerobic conditions using glycerol as the sole carbon source and in the absence of other exogenous reducing equivalent acceptors. Under these conditions, in for example, strains of *Citrobacter*, *Clostridium*, and *Klebsiella*, a parallel pathway for glycerol operates which first involves oxidation of glycerol to dihydroxyacetone (DHA) by a NAD<sup>+</sup>- (or NADP<sup>+</sup>-) linked glycerol dehydrogenase (Equation 3). The DHA, following phosphorylation to dihydroxyacetone phosphate (DHAP) by a DHA kinase (Equation 4), becomes available for biosynthesis and for supporting ATP generation via, for example, glycolysis.



In contrast to the 1,3-propanediol pathway, this pathway may provide carbon and energy to the cell and produces rather than consumes NADH.

In *Klebsiella pneumoniae* and *Citrobacter freundii*, the genes encoding the functionally linked activities of glycerol dehydratase (*dhaB*), 1,3-propanediol oxidoreductase (*dhaT*), glycerol dehydrogenase (*dhaD*), and dihydroxyacetone kinase (*dhaK*) are encompassed by the *dha* regulon. The *dha* regulons from *Citrobacter* and *Klebsiella* have been expressed in *Escherichia coli* and have been shown to convert glycerol to 1,3-propanediol. Glycerol dehydratase (E.C. 4.2.1.30) and diol [1,2-propanediol] dehydratase (E.C. 4.2.1.28) are related but distinct enzymes that are encoded by distinct genes. In *Salmonella typhimurium* and *Klebsiella pneumoniae*, diol dehydratase is associated with the *pdu* operon, see Bobik et al., 1992, J. Bacteriol. 174:2253-2266 and United States patent 5,633,362. Tobimatsu, et al., 1996, J. Biol. Chem. 271: 22352-22357 disclose the *K. pneumoniae* gene encoding glycerol dehydratase protein X identified as ORF 4; Segfried et al., 1996, J. Bacteriol. 178: 5793-5796 disclose the *C. freundii* glycerol dehydratase gene encoding protein X identified as ORF Z. Tobimatsu et al., 1995, J. Biol. Chem. 270:7142-7148 disclose the diol dehydratase submits  $\alpha$ ,  $\beta$  and  $\gamma$  and illustrate the presence of orf 4. Luers (1997, FEMS Microbiology Letters 154:337-345) disclose the amino acid sequence of protein 1, protein 2 and protein 3 of *Clostridium pasteurianum*.

Biological processes for the preparation of glycerol are known. The overwhelming majority of glycerol producers are yeasts, but some bacteria, other fungi and algae are also known to produce glycerol. Both bacteria and yeasts produce glycerol by converting glucose or other carbohydrates through the fructose-1,6-bisphosphate pathway in glycolysis or by the Embden Meyerhof Parnas pathway, whereas, certain algae convert dissolved carbon dioxide or bicarbonate in the chloroplasts into the 3-carbon intermediates of the Calvin cycle. In a series of steps, the 3-carbon intermediate, phosphoglyceric acid, is converted to glyceraldehyde 3-phosphate which can be readily interconverted to its keto isomer dihydroxyacetone phosphate and ultimately to glycerol.



-- 3 --

Specifically, the bacteria *Bacillus licheniformis* and *Lactobacillus lycopersica* synthesize glycerol, and glycerol production is found in the halotolerant algae *Dunaliella sp.* and *Asteromonas gracilis* for protection against high external salt concentrations (Ben-Amotz et al., *Experientia* 38, 49-52, (1982)). Similarly, various osmotolerant yeasts synthesize glycerol as a protective  
5 measure. Most strains of *Saccharomyces* produce some glycerol during alcoholic fermentation, and this can be increased physiologically by the application of osmotic stress (Albertyn et al., *Mol. Cell. Biol.* 14, 4135-4144, (1994)). Earlier this century commercial glycerol production was achieved by the use of *Saccharomyces* cultures to which "steering reagents" were added such as sulfites or alkalis. Through the formation of an inactive complex, the steering agents block or  
10 inhibit the conversion of acetaldehyde to ethanol; thus, excess reducing equivalents (NADH) are available to or "steered" towards DHAP for reduction to produce glycerol. This method is limited by the partial inhibition of yeast growth that is due to the sulfites. This limitation can be partially overcome by the use of alkalis which create excess NADH equivalents by a different mechanism. In this practice, the alkalis initiated a Cannizzaro disproportionation to yield ethanol and acetic acid  
15 from two equivalents of acetaldehyde.

The gene encoding glycerol-3-phosphate dehydrogenase (DAR1, GPD1) has been cloned and sequenced from *S. diastaticus* (Wang et al., *J. Bact.* 176, 7091-7095, (1994)). The DAR1 gene was cloned into a shuttle vector and used to transform *E. coli* where expression produced active enzyme. Wang et al. (supra) recognize that DAR1 is regulated by the cellular osmotic  
20 environment but do not suggest how the gene might be used to enhance 1,3-propanediol production in a recombinant organism.

Other glycerol-3-phosphate dehydrogenase enzymes have been isolated: for example, sn-glycerol-3-phosphate dehydrogenase has been cloned and sequenced from *S. cerevisiae* (Larason et al., *Mol. Microbiol.* 10, 1101, (1993)) and Albertyn et al., (*Mol. Cell. Biol.* 14, 4135,  
25 (1994)) teach the cloning of GPD1 encoding a glycerol-3-phosphate dehydrogenase from *S. cerevisiae*. Like Wang et al. (supra), both Albertyn et al. and Larason et al. recognize the osmo-sensitivity of the regulation of this gene but do not suggest how the gene might be used in the production of 1,3-propanediol in a recombinant organism.

As with G3PDH, glycerol-3-phosphatase has been isolated from *Saccharomyces cerevisiae* and the protein identified as being encoded by the GPP1 and GPP2 genes (Norbeck et al., *J. Biol. Chem.* 271, 13875, (1996)). Like the genes encoding G3PDH, it appears that GPP2 is osmosensitive.

Although biological methods of both glycerol and 1,3-propanediol production are known, it has never been demonstrated that the entire process can be accomplished by a single  
35 recombinant organism.

Neither the chemical nor biological methods described above for the production of 1,3-propanediol are well suited for industrial scale production since the chemical processes are energy intensive and the biological processes require the expensive starting material, glycerol. A

method requiring low energy input and an inexpensive starting material is needed. A more desirable process would incorporate a microorganism that would have the ability to convert basic carbon sources such as carbohydrates or sugars to the desired 1,3-propanediol end-product.

Although a single organism conversion of fermentable carbon source other than glycerol or dihydroxyacetone to 1,3-propanediol would be desirable, it has been documented that there are significant difficulties to overcome in such an endeavor. For example, Gottschalk et al. (EP 373 230) teach that the growth of most strains useful for the production of 1,3-propanediol, including *Citrobacter freundii*, *Clostridium autobutylicum*, *Clostridium butylicum*, and *Klebsiella pneumoniae*, is disturbed by the presence of a hydrogen donor such as fructose or glucose. Strains of *Lactobacillus brevis* and *Lactobacillus buchneri*, which produce 1,3-propanediol in co-fermentations of glycerol and fructose or glucose, do not grow when glycerol is provided as the sole carbon source, and, although it has been shown that resting cells can metabolize glucose or fructose, they do not produce 1,3-propanediol. (Veiga DA Cunha et al., *J. Bacteriol.* 174, 1013 (1992)). Similarly, it has been shown that a strain of *Llyobacter polytropus*, which produces 1,3-propanediol when glycerol and acetate are provided, will not produce 1,3-propanediol from carbon substrates other than glycerol, including fructose and glucose. (Steib et al., *Arch. Microbiol.* 140, 139 (1984)). Finally Tong et al. (*Appl. Biochem. Biotech.* 34, 149 (1992)) has taught that recombinant *Escherichia coli* transformed with the *dha* regulon encoding glycerol dehydratase does not produce 1,3-propanediol from either glucose or xylose in the absence of exogenous glycerol.

Attempts to improve the yield of 1,3-propanediol from glycerol have been reported where co-substrates capable of providing reducing equivalents, typically fermentable sugars, are included in the process. Improvements in yield have been claimed for resting cells of *Citrobacter freundii* and *Klebsiella pneumoniae* DSM 4270 cofermenting glycerol and glucose (Gottschalk et al., *supra.*, and Tran-Dinh et al., DE 3734 764); but not for growing cells of *Klebsiella pneumoniae* ATCC 25955 cofermenting glycerol and glucose, which produced no 1,3-propanediol (I-T. Tong, Ph.D. Thesis, University of Wisconsin-Madison (1992)). Increased yields have been reported for the cofermentation of glycerol and glucose or fructose by a recombinant *Escherichia coli*; however, no 1,3-propanediol is produced in the absence of glycerol (Tong et al., *supra.*). In these systems, single organisms use the carbohydrate as a source of generating NADH while providing energy and carbon for cell maintenance or growth. These disclosures suggest that sugars do not enter the carbon stream that produces 1,3-propanediol. In no case is 1,3-propanediol produced in the absence of an exogenous source of glycerol. Thus the weight of literature clearly suggests that the production of 1,3-propanediol from a carbohydrate source by a single organism is not possible.

-- 5 --

The weight of literature regarding the role of protein X in 1,3-propanediol production by a host cell is at best confusing. Prior to the availability of gene information, McGee et al., 1982, Biochem. Biophys. Res. Comm. 108: 547-551, reported diol dehydratase from *K. pneumoniae* ATCC 8724 to be composed of four subunits identified by size (60K, 51K, 29K, and 15K daltons) and N-terminal amino acid sequence. In direct contrast to McGee, Tobimatsu et al. 1995, *supra*, report the cloning, sequencing and expression of diol dehydratase from the same organism and find no evidence linking the 51K dalton polypeptide to dehydratase. Tobimatsu et al. 1996, *supra*, conclude that the protein X polypeptide is not a subunit of glycerol dehydratase, in contrast to GenBank Accession Number U30903 where protein X is described as a large subunit of glycerol dehydratase. Seyfried et al., *supra*, report that a deletion of 192 bp from the 3' end of *orfZ* (protein X) had no effect on enzyme activity and conclude that *orfZ* does not encode a subunit required for dehydratase activity. Finally, Skraly, F.A. (1997, Thesis entitled "Metabolic Engineering of an Improved 1,3-Propanediol Fermentation") disclose a loss of glycerol dehydratase activity in one experiment where recombinant ORF3 (proteinX) was disrupted creating a large fusion protein but not in another experiment where 1,3-propanediol production from glycerol was diminished compared to a control where ORF3 was intact.

The problem to be solved by the present invention is the biological production of 1,3-propanediol by a single recombinant organism from an inexpensive carbon substrate such as glucose or other sugars in commercially feasible quantities. The biological production of 1,3-propanediol requires glycerol as a substrate for a two step sequential reaction in which a dehydratase enzyme (typically a coenzyme B<sub>12</sub>-dependent dehydratase) converts glycerol to an intermediate, 3-hydroxypropionaldehyde, which is then reduced to 1,3-propanediol by a NADH- (or NADPH) dependent oxidoreductase. The complexity of the cofactor requirements necessitates the use of a whole cell catalyst for an industrial process which utilizes this reaction sequence for the production of 1,3-propanediol. Furthermore, in order to make the process economically viable, a less expensive feedstock than glycerol or dihydroxyacetone is needed and high production levels are desirable. Glucose and other carbohydrates are suitable substrates, but, as discussed above, are known to interfere with 1,3-propanediol production. As a result no single organism has been shown to convert glucose to 1,3-propanediol.

Applicants have solved the stated problem and the present invention provides for bioconverting a fermentable carbon source directly to 1,3-propanediol using a single recombinant organism. Glucose is used as a model substrate and the bioconversion is applicable to any existing microorganism. Microorganisms harboring the genes encoding protein X and protein 1, protein 2 and protein 3 in addition to other proteins associated with the production of 1,3-propanediol, are able to convert glucose and other sugars through the glycerol degradation pathway to 1,3-propanediol with good yields and selectivities. Furthermore, the present invention

-- 6 --

may be generally applied to include any carbon substrate that is readily converted to 1) glycerol, 2) dihydroxyacetone, or 3) C<sub>3</sub> compounds at the oxidation state of glycerol (e.g., glycerol 3-phosphate) or 4) C<sub>3</sub> compounds at the oxidation state of dihydroxyacetone (e.g., dihydroxyacetone phosphate or glyceraldehyde 3-phosphate).

5 **Summary of the Invention**

The present invention relates to improved methods for the production of 1,3-propanediol from a single microorganism. The present invention is based, in part, upon the unexpected discovery that the presence of a gene encoding protein X in a microorganism containing at least one gene encoding a dehydratase activity and capable of producing 1,3-propanediol is associated with the *in vivo* reactivation of dehydratase activity and increased production of 1,3-propanediol in the microorganism. The present invention is also based, in part, upon the unexpected discovery that the presence of a gene encoding protein X and at least one gene encoding a protein selected from the group consisting of protein 1, protein 2 and protein 3 in host cells containing at least one gene encoding a dehydratase activity and capable of producing 1,3-propanediol is associated with *in vivo* reactivation of the dehydratase activity and increased yields of 1,3-propanediol in the microorganism.

Accordingly, the present invention provides an improved method for the production of 1,3-propanediol from a microorganism capable of producing 1,3-propanediol, said microorganism comprising at least one gene encoding a dehydratase activity, the method comprising the steps of introducing a gene encoding protein X into the organism to create a transformed organism; and culturing the transformed organism in the presence of at least one carbon source capable of being converted to 1,3 propanediol in said transformed host organism and under conditions suitable for the production of 1,3 propanediol wherein the carbon source is selected from the group consisting of monosaccharides, oligosaccharides, polysaccharides, and a one carbon substrate.

In a preferred embodiment, the method for improved production of 1,3-propanediol further comprises introducing at least one gene encoding a protein selected from the group consisting of protein 1, protein 2 and protein 3 into the organism. The microorganism may further comprise at least one of (a) a gene encoding a glycerol-3-phosphate dehydrogenase activity; (b) a gene encoding a glycerol-3-phosphatase activity; and (c) a gene encoding 1,3-propanediol oxidoreductase activity into the microorganism. Gene(s) encoding a dehydratase activity, protein X, proteins 1, 2 or 3 or other genes necessary for the production of 1,3-propanediol may be stably maintained in the host cell genome or may be on replicating plasmids residing in the host microorganism.

35 The method optionally comprises the step of recovering the 1,3 propanediol. In one aspect of the present invention, the carbon source is glucose.

-- 7 --

The microorganism is selected from the group of genera consisting of *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*, *Bacillus*,  
5 *Streptomyces* and *Pseudomonas*.

In one aspect, protein X is derived from a glyceol dehydratase gene cluster and in another aspect, protein X is derived from a diol dehydratase gene cluster. The gene encoding the dehydratase activity may be homologous to the microorganism or heterologous to the microorganism. In one embodiment, the glycerol dehydratase gene cluster is derived from an  
10 organism selected from the genera consisting of *Klebsiella* and *Citrobacter*. In another embodiment, the diol dehydratase gene cluster is derived from an organism selected from the genera consisting of *Klebsiella*, *Clostridium* and *Salmonella*.

In another aspect, the present invention provides a recombinant microorganism comprising at least one gene encoding a dehydratase activity; at least one gene encoding a  
15 glycerol-3-phosphatase; and at least one gene encoding protein X, wherein said microorganism is capable of producing 1,3-propanediol from a carbon source. The carbon source may be selected from the group consisting of monosaccharides, oligosaccharides, polysaccharides, and a one carbon substrate. In a further embodiment, the microorganism further comprises a gene encoding a cytosolic glycerol-3-phosphate dehydrogenase. In another embodiment, the  
20 recombinant microorganism further comprises at least one gene encoding a protein selected from the group consisting of protein 1, protein 2 and protein 3. The microorganism is selected from the group consisting of *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*,  
25 *Escherichia*, *Salmonella*, *Bacillus*, *Streptomyces* and *Pseudomonas*. In one aspect, protein X is derived from a glycerol dehydratase gene cluster. In another aspect, protein X is derived from a diol dehydratase gene cluster. In one aspect, the dehydratase activity is heterologous to said microorganism and in another aspect, the dehydratase activity is homologous to said microorganism.

30 The present invention also provides a method for the *in vivo* reactivation of a dehydratase activity in a microorganism capable of producing 1,3-propanediol and containing at least one gene encoding a dehydratase activity, comprising the step of introducing a gene encoding protein X into said microorganism. The microorganism is selected from the group consisting of *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*,  
35 *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*,

*Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*, *Bacillus*, *Streptomyces* and *Pseudomonas*.

In one aspect, the gene encoding the dehydratase activity is heterologous to said microorganism and in another aspect, the gene encoding the dehydratase activity is homologous to said microorganism. In one embodiment, the gene encoding protein X is derived from a glycerol dehydratase gene cluster and in another embodiment, the gene encoding protein X is derived from a diol dehydratase gene cluster.

The present invention also provides expression vectors and host cells containing genes encoding protein X, protein 1, protein 2 and protein 3.

One advantage of the method of production of 1,3-propanediol according to the present invention is the unexpected increased production of 1,3-propanediol in a host cell capable of producing 1,3-propanediol in the presence of nucleic acid encoding protein X as compared to the host cell lacking nucleic acid encoding protein X. As demonstrated *infra*, a host cell containing nucleic acid encoding dhaB 1, 2 and 3 and protein X is able to produce significantly more 1,3-propanediol than a host cell containing nucleic acid encoding dhaB 1, 2 and 3 and lacking X.

Another advantage of the present invention as demonstrated *infra*, is that the presence of nucleic acid encoding protein X along with nucleic acid encoding at least one of protein 1, protein 2 and protein 3 in a host cell capable of producing 1,3-propanediol gives the unexpected result of increased production of 1,3-propanediol in the host cell over 1,3-propanediol production in the host cell lacking nucleic acid encoding protein X along with nucleic acid encoding at least one of protein 1, protein 2 and protein 3.

Yet another advantage of the method of production of the present invention as shown *infra* is the *in vivo* reactivation of the dehydratase activity in a microorganism that is associated with the presence of nucleic acid encoding protein X in the microorganism.

#### **Brief Description of the Drawings**

Figure 1 illustrates components of the glycerol dehydratase gene cluster from *Klebsiella pneumoniae* on plasmid pHK28-26 (SEQ ID NO:19). In this figure, orfY encodes protein 1, orfX encodes protein 2 and orfW encodes protein 3. DhaB-X refers to protein X.

Figures 2A-2G illustrates the nucleotide and amino acid sequence of *Klebsiella pneumoniae* glycerol dehydratase protein X (dhab4) (SEQ ID NO:59).

Figure 3 illustrates the amino acid alignment of *Klebsiella pneumoniae* protein 1 (SEQ ID NO: 61) and *Citrobacter freundii* protein1 (SEQ ID NO: 60) (designated in Figure 3 as orfY).

Figure 4 illustrates the amino acid alignment of *Klebsiella pneumoniae* protein 2 (SEQ ID NO: 63) and *Citrobacter freundii* protein 2 (SEQ ID NO: 62) (designated in Figure 4 as orfX).

Figure 5 illustrates the amino acid alignment of *Klebsiella pneumoniae* protein 3 (SEQ ID NO: 64) and *Citrobacter freundii* protein 3 (SEQ ID NO: 65) (designated in Figure 5 as orfW).

-- 9 --

Figure 6 illustrates the in situ reactivation comparison of plasmids pHK28-26 (which contains *dhaB* subunits 1, 2 and 3 as well as protein X and the open reading frames encoding protein 1, protein 2 and protein 3) vs. pDT24 (which contains *dhaB* subunits 1, 2 and 3 as well as protein X) in *E.coli* DH5 $\alpha$  cells.

5 Figure 7 illustrate the in situ reactivation comparison of plasmids pM7 (containing genes encoding *dhaB* subunits 1, 2 and 3 and protein X) vs. Plasmid pM11 (containing genes encoding *dhaB* subunits 1, 2 and 3) in *E.coli* DH5 $\alpha$  cells.

Figures 8A-8E illustrates the nucleic acid (SEQ ID NO: 66) and amino acid (SEQ ID NO: 67) sequence of *K. pneumoniae* diol dehydratase gene cluster protein X.

10 Figure 9 illustrates a standard 10 liter fermentation for 1,3 propandiol production using *E. coli* FM5/pDT24 (FM5 described in Amgen patent US 5,494,816 , ATCC accession No. 53911).

Figure 10 illustrates a standard 10 liter fermentation for 1,3 propandiol production using *E. coli* DH5alpha/pHK28-26.

15 **Brief Description of Biological Deposits and Sequence Listing**

The transformed *E. coli* W2042 (comprising the *E. coli* host W1485 and plasmids pDT20 and pAH42) containing the genes encoding glycerol-3-phosphate dehydrogenase (G3PDH) and glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*) was deposited on 26 September 1996 with the ATCC under the terms of  
20 the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for the Purpose of Patent Procedure and is designated as ATCC 98188.

*S. cerevisiae* YPH500 harboring plasmids pMCK10, pMCK17, pMCK30 and pMCK35 containing genes encoding glycerol-3-phosphate dehydrogenase (G3PDH) and glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol  
25 oxidoreductase (*dhaT*) was deposited on 26 September 1996 with the ATCC under the terms of the Budapest Treaty on the International Recognition of the Deposit of Micro-organisms for the Purpose of Patent Procedure and is designated as ATCC 74392.

*E.coli* DH5 $\alpha$  containing pKP1 which has about 35kb of a *Klebsiella* genome which contains the glycerol dehydratase, protein X and proteins 1, 2 and 3 was deposited on 18 April 1995 with  
30 the ATCC under the terms of the Budapest Treaty and was designated ATCC 69789. *E.coli* DH5 $\alpha$  containing pKP4 containing a portion of the *Klebsiella* genome encoding diol dehydratase enzyme, including protein X was deposited on 18 April 1995 with the ATCC under the terms of the Budapest Treaty and was designated ATCC 69790.

"ATCC" refers to the American Type Culture Collection international depository located at  
35 12301 Parklawn Drive, Rockville, MD 20852 U.S.A. The designations refer to the accession number of the deposited material.

-- 10 --

**Detailed Description of the Invention**

The present invention relates to the production of 1,3-propanediol in a single microorganism and provides improved methods for production of 1,3-propanediol from a fermentable carbon source in a single recombinant organism. The method incorporates a microorganism capable of producing 1,3-propanediol comprising either homologous or heterologous genes encoding dehydratase (*dhaB*), at least one gene encoding protein X and optionally at least one of the genes encoding a protein selected from the group consisting of protein 1, protein 2 and protein 3. Optionally, the microorganism contains at least one gene encoding glycerol-3-phosphate dehydrogenase, glycerol-3-phosphatase and 1,3-propanediol oxidoreductase (*dhaT*). The recombinant microorganism is contacted with a carbon substrate and 1,3-propanediol is isolated from the growth media.

The present method provides a rapid, inexpensive and environmentally responsible source of 1,3-propanediol monomer useful in the production of polyesters and other polymers.

The following definitions are to be used to interpret the claims and specification.

The term "dehydratase gene cluster" or "gene cluster" refers to the set of genes which are associated with 1,3-propanediol production in a host cell and is intended to encompass glycerol dehydratase gene clusters as well as diol dehydratase gene clusters. The *dha* regulon refers to a glycerol dehydratase gene cluster, as illustrated in Figure 1 which includes regulatory regions.

The term "regenerating the dehydratase activity" or "reactivating the dehydratase activity" refers to the phenomenon of converting a dehydratase not capable of catalysis of a substrate to one capable of catalysis of a substrate or to the phenomenon of inhibiting the inactivation of a dehydratase or the phenomenon of extending the useful half-life of the dehydratase enzyme *in vivo*.

The terms "glycerol dehydratase" or "dehydratase enzyme" or "dehydratase activity" refer to the polypeptide(s) responsible for an enzyme activity that is capable of isomerizing or converting a glycerol molecule to the product 3-hydroxypropionaldehyde. For the purposes of the present invention the dehydratase enzymes include a glycerol dehydratase (GenBank U09771, U30903) and a diol dehydratase (GenBank D45071) having preferred substrates of glycerol and 1,2-propanediol, respectively. Glycerol dehydratase of *K. pneumoniae* ATCC 25955 is encoded by the genes *dhaB1*, *dhaB2*, and *dhaB3* identified as SEQ ID NOS:1, 2 and 3, respectively. The *dhaB1*, *dhaB2*, and *dhaB3* genes code for the a, b, and c subunits of the glycerol dehydratase enzyme, respectively.

The phrase "protein X of a dehydratase gene cluster" or "dhaB protein X" or "protein X" refers to a protein that is comparable to protein X of the *Klebsiella pneumoniae* dehydratase gene cluster as shown in Figure 2 or alternatively comparable to protein X of *Klebsiella pneumoniae* diol dehydratase gene cluster as shown in Figure 8. Preferably protein X is capable of increasing



-- 11 --

the production of 1,3-propanediol in a host organism over the production of 1,3-propanediol in the absence of protein X in the host organism. Being comparable means that DNA encoding the protein is either in the same structural location as DNA encoding *Klebsiella* protein X with respect to *Klebsiella* dhaB1, dhaB2 and dhaB3, i.e., DNA encoding protein X is 3' to nucleic acid encoding dhaB1-B3, or that protein X has overall amino acid similarity to either *Klebsiella* diol or glycerol dehydratase protein X. The present invention encompasses protein X molecules having at least 50%; or at least 65 %; or at least 80%; or at least 90% or at least 95% similarity to the protein X of *K. pneumoniae* glycerol or diol dehydratase or the *C. freundii* protein X.

Included within the term "protein X" is protein X, also referred to as ORF Z, from *Citrobacter* dha regulon (Segfried M. 1996, J. Bacteriol. 178: 5793:5796). The present invention also encompasses amino acid variations of protein X from any microorganism as long as the protein X variant retains its essential functional characteristics of increasing the production of 1,3-propanediol in a host organism over the production of 1,3-propanediol in the host organism in the absence of protein X.

A portion of the *Klebsiella* genome encoding the glycerol dehydratase enzyme activity as well as protein X was transformed into *E.coli* and the transformed *E.coli* was deposited on 18 April 1995 with the ATCC under the terms of the Budapest Treaty and was designated as ATCC accession number 69789. A portion of the *Klebsiella* genome encoding the diol dehydratase enzyme activity as well as protein X was transformed into *E.coli* and the transformed *E.coli* was deposited on 18 April 1995 with the ATCC under the terms of the Budapest Treaty and was designated as ATCC accession number 69790.

*Klebsiella* glycerol dehydratase protein X is found at bases 9749-11572 of SEQ ID NO:19, counting the first base of dhaK as position number 1. *Citrobacter freundii* (ATCC accession number CFU09771) nucleic acid encoding protein X is found between positions 11261 and 13072.

The present invention encompasses genes encoding dehydratase protein X that are recombinantly introduced and replicate on a plasmid in the host organism as well as genes that are stably maintained in the host genome. The present invention encompasses a method for enhanced production of 1,3-propanediol wherein the gene encoding protein X is transformed in a host cell together with genes encoding the dehydratase activity and/or other genes necessary for the production of 1,3-propanediol. The gene encoding protein X, dehydratase activity and/or other genes may be on the same or different expression cassettes. Alternatively, the gene encoding protein X may be transformed separately, either before or after genes encoding the dehydratase activity and/or other activities. The present invention encompasses host cell having endogenous nucleic acid encoding protein X as well as host cell lacking endogenous nucleic acid encoding protein X.

-- 12 --

The terms "protein 1", "protein 2" and "protein 3" refer to the proteins encoded in a microorganism that are comparable to protein 1 (SEQ ID NO: 60 or SEQ ID NO: 61)(also referred to as orfY), protein 2 (SEQ ID NO: 62 or SEQ ID NO: 63) (also referred to as orfX) and protein 3 (SEQ ID NO: 64 or SEQ ID NO: 65) (also referred to as orfW), respectively.

5 Preferably, in the presence of protein X, at least one of proteins 1, 2 and 3 is capable of increasing the production of 1,3-propanediol in a host organism over the production of 1,3-propanediol in the absence of protein X and at least one of proteins 1, 2 and 3 in the host organism. Being comparable means that DNA encoding the protein is either in the same structural location as DNA encoding the respective proteins, as shown in Figure 1, or that the  
10 respective proteins have overall amino acid similarity to the respective SEQ ID NOS shown in Figures 3, 4 and 5.

The present invention encompasses protein 1 molecules having at least 50%; or at least 65 %; or at least 80%; or at least 90% or at least 95% similarity to SEQ ID NO: 60 or SEQ ID NO: 61. The present invention encompasses protein 2 molecules having at least 50%; or at least 65  
15 %; or at least 80%; or at least 90% or at least 95% similarity to SEQ ID NO: 62 or SEQ ID NO: 63. The present invention encompasses protein 3 molecules having at least 50%; or at least 65 %; or at least 80%; or at least 90% or at least 95% similarity to SEQ ID NO: 64 or SEQ ID NO: 65.

Included within the terms "protein 1", "protein 2" and "protein 3", respectively, are orfY, orfX and orfW from *Clostridium pasteurianum* (Luers, et al., *supra*) as well as molecules having at  
20 least 50%; or at least 65 %; or at least 80%; or at least 90% or at least 95% similarity to C. *pasterurianum* orfY, orfX or orfW. The present invention also encompasses amino acid variations of proteins 1, 2 and 3 from any microorganism as long as the protein variant, in combination with protein X, retains its essential functional characteristics of increasing the production of 1,3-propanediol in a host organism over the production of 1,3-propanediol in the host organism in  
25 their absence.

The present invention encompasses a method for enhanced production of 1,3-propanediol wherein the gene(s) encoding at least one of protein 1, protein 2 and protein 3 is transformed in a host cell together with genes encoding protein X, the dehydratase activity and/or other genes necessary for the production of 1,3-propanediol. The gene(s) encoding at least on of proteins 1, 2  
30 and 3, protein X, dehydratase activity and/or other genes may be on the same or different expression cassettes. Alternatively, the gene(s) encoding at least one of proteins 1, 2 and 3 may be transformed separately, either before or after genes encoding the dehydratase activity and/or other activities. The present invention encompasses host cell having endogenous nucleic acid encoding protein 1, protein 2 or protein 3 as well as host cell lacking endogenous nucleic acid  
35 encoding the proteins.

-- 13 --

The terms "oxidoreductase" or "1,3-propanediol oxidoreductase" refer to the polypeptide(s) responsible for an enzyme activity that is capable of catalyzing the reduction of 3-hydroxypropionaldehyde to 1,3-propanediol. 1,3-Propanediol oxidoreductase includes, for example, the polypeptide encoded by the *dhaT* gene (GenBank U09771, U30903) and is identified  
5 as SEQ ID NO:4.

The terms "glycerol-3-phosphate dehydrogenase" or "G3PDH" refer to the polypeptide(s) responsible for an enzyme activity capable of catalyzing the conversion of dihydroxyacetone phosphate (DHAP) to glycerol-3-phosphate (G3P). *In vivo* G3PDH may be NADH-, NADPH-, or FAD-dependent. Examples of this enzyme activity include the following: NADH-dependent  
10 enzymes (EC 1.1.1.8) are encoded by several genes including GPD1 (GenBank Z74071x2) or GPD2 (GenBank Z35169x1) or GPD3 (GenBank G984182) or DAR1 (GenBank Z74071x2); a NADPH-dependent enzyme (EC 1.1.1.94) is encoded by *gpsA* (GenBank U32164, G466746 (cds 197911-196892), and L45246); and FAD-dependent enzymes (EC 1.1.99.5) are encoded by GUT2 (GenBank Z47047x23) or glpD (GenBank G147838) or glpABC (GenBank M20938).

15 The terms "glycerol-3-phosphatase" or "sn-glycerol-3-phosphatase" or "d,l-glycerol phosphatase" or "G3P phosphatase" refer to the polypeptide(s) responsible for an enzyme activity that is capable of catalyzing the conversion of glycerol-3-phosphate to glycerol. G3P phosphatase includes, for example, the polypeptides encoded by GPP1 (GenBank Z47047x125) or GPP2 (GenBank U18813x11).

20 The term "glycerol kinase" refers to the polypeptide(s) responsible for an enzyme activity capable of catalyzing the conversion of glycerol to glycerol-3-phosphate or glycerol-3-phosphate to glycerol, depending on reaction conditions. Glycerol kinase includes, for example, the polypeptide encoded by GUT1 (GenBank U11583x19).

The terms "GPD1", "DAR1", "OSG1", "D2830", and "YDL022W" will be used  
25 interchangeably and refer to a gene that encodes a cytosolic glycerol-3-phosphate dehydrogenase and characterized by the base sequence given as SEQ ID NO:5.

The term "GPD2" refers to a gene that encodes a cytosolic glycerol-3-phosphate dehydrogenase and characterized by the base sequence given as SEQ ID NO:6.

30 The terms "GUT2" and "YIL155C" are used interchangeably and refer to a gene that encodes a mitochondrial glycerol-3-phosphate dehydrogenase and characterized by the base sequence given in SEQ ID NO:7.

The terms "GPP1", "RHR2" and "YIL053W" are used interchangeably and refer to a gene that encodes a cytosolic glycerol-3-phosphatase and characterized by the base sequence given as SEQ ID NO:8.

-- 14 --

The terms "GPP2", "HQR2" and "YER062C" are used interchangeably and refer to a gene that encodes a cytosolic glycerol-3-phosphatase and characterized by the base sequence given as SEQ ID NO:9.

5 The term "GUT1" refers to a gene that encodes a cytosolic glycerol kinase and characterized by the base sequence given as SEQ ID NO:10.

The terms "function" or "enzyme function" refer to the catalytic activity of an enzyme in altering the energy required to perform a specific chemical reaction. It is understood that such an activity may apply to a reaction in equilibrium where the production of either product or substrate may be accomplished under suitable conditions.

10 The terms "polypeptide" and "protein" are used interchangeably.

The terms "carbon substrate" and "carbon source" refer to a carbon source capable of being metabolized by host organisms of the present invention and particularly carbon sources selected from the group consisting of monosaccharides, oligosaccharides, polysaccharides, and one-carbon substrates or mixtures thereof.

15 The terms "host cell" or "host organism" refer to a microorganism capable of receiving foreign or heterologous genes and of expressing those genes to produce an active gene product.

The terms "foreign gene", "foreign DNA", "heterologous gene" and "heterologous DNA" refer to genetic material native to one organism that has been placed within a host organism by various means. The gene of interest may be a naturally occurring gene, a mutated gene or a  
20 synthetic gene.

The terms "recombinant organism" and "transformed host" refer to any organism having been transformed with heterologous or foreign genes or extra copies of homologous genes. The recombinant organisms of the present invention express foreign genes encoding glycerol-3-phosphate dehydrogenase (G3PDH) and glycerol-3-phosphatase (G3P phosphatase), glycerol  
25 dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*) for the production of 1,3-propanediol from suitable carbon substrates.

"Gene" refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding) and following (3' non-coding) the coding region. The terms "native" and "wild-type" refer to a gene as found in nature with its own regulatory  
30 sequences.

The terms "encoding" and "coding" refer to the process by which a gene, through the mechanisms of transcription and translation, produces an amino acid sequence. It is understood that the process of encoding a specific amino acid sequence includes DNA sequences that may involve base changes that do not cause a change in the encoded amino acid, or which involve  
35 base changes which may alter one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. It is therefore understood that the

-- 15 --

invention encompasses more than the specific exemplary sequences. Modifications to the sequence, such as deletions, insertions, or substitutions in the sequence which produce silent changes that do not substantially affect the functional properties of the resulting protein molecule are also contemplated. For example, alteration in the gene sequence which reflect the  
5 degeneracy of the genetic code, or which result in the production of a chemically equivalent amino acid at a given site, are contemplated. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result in substitution of one negatively charged residue for another, such as aspartic acid  
10 for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a biologically equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. In some cases, it may in fact be desirable to make mutants of the sequence in order to study the effect of alteration on the biological activity of the  
15 protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity in the encoded products. Moreover, the skilled artisan recognizes that sequences encompassed by this invention are also defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65 °C), with the sequences exemplified herein.

20 The term "expression" refers to the transcription and translation to gene product from a gene coding for the sequence of the gene product.

The terms "plasmid", "vector", and "cassette" refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements may be autonomously  
25 replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. "Transformation cassette"  
30 refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitate transformation of a particular host cell. "Expression cassette" refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that allow for enhanced expression of that gene in a foreign host.

The terms "transformation" and "transfection" refer to the acquisition of new genes in a  
35 cell after the incorporation of nucleic acid. The acquired genes may be integrated into

-- 16 --

chromosomal DNA or introduced as extrachromosomal replicating sequences. The term "transformant" refers to the product of a transformation.

The term "genetically altered" refers to the process of changing hereditary material by transformation or mutation.

5 The term "isolated" refers to a protein or DNA sequence that is removed from at least one component with which it is naturally associated.

The term "homologous" refers to a protein or polypeptide native or naturally occurring in a gram-positive host cell. The invention includes microorganisms producing the homologous protein via recombinant DNA technology.

10

### CONSTRUCTION OF RECOMBINANT ORGANISMS

Recombinant organisms containing the necessary genes that will encode the enzymatic pathway for the conversion of a carbon substrate to 1,3-propanediol may be constructed using techniques well known in the art. As discussed in Example 9, genes encoding *Klebsiella* dhaB1, dhaB2, dhaB3 and protein X were used to transform *E. coli* DH5a and in Example 10, genes encoding at least one of *Klebsiella* proteins 1, 2 and 3 as well as at least one gene encoding protein X was used to transform *E. coli*.

Genes encoding glycerol-3-phosphate dehydrogenase (G3PDH), glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*) were isolated from a native host such as *Klebsiella* or *Saccharomyces* and used to transform host strains such as *E. coli* DH5a, ECL707, AA200, or W1485; the *Saccharomyces cerevisiae* strain YPH500; or the *Klebsiella pneumoniae* strains ATCC 25955 or ECL 2106.

#### Isolation of Genes

Methods of obtaining desired genes from a bacterial genome are common and well known in the art of molecular biology. For example, if the sequence of the gene is known, suitable genomic libraries may be created by restriction endonuclease digestion and may be screened with probes complementary to the desired gene sequence. Once the sequence is isolated, the DNA may be amplified using standard primer directed amplification methods such as polymerase chain reaction (PCR) (U.S. 4,683,202) to obtain amounts of DNA suitable for transformation using appropriate vectors.

Alternatively, cosmid libraries may be created where large segments of genomic DNA (35-45kb) may be packaged into vectors and used to transform appropriate hosts. Cosmid vectors are unique in being able to accommodate large quantities of DNA. Generally, cosmid vectors have at least one copy of the *cos* DNA sequence which is needed for packaging and subsequent circularization of the foreign DNA. In addition to the *cos* sequence these vectors will also contain an origin of replication such as ColE1 and drug resistance markers such as a gene resistant to ampicillin or neomycin. Methods of using cosmid vectors for the transformation of

35

-- 17 --

suitable bacterial hosts are well described in Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Edition (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

Typically to clone cosmids, foreign DNA is isolated and ligated, using the appropriate restriction endonucleases, adjacent to the *cos* region of the cosmid vector. Cosmid vectors containing the linearized foreign DNA is then reacted with a DNA packaging vehicle such as bacteriophage  $\lambda$ . During the packaging process the *cos* sites are cleaved and the foreign DNA is packaged into the head portion of the bacterial viral particle. These particles are then used to transfect suitable host cells such as *E. coli*. Once injected into the cell, the foreign DNA circularizes under the influence of the *cos* sticky ends. In this manner large segments of foreign DNA can be introduced and expressed in recombinant host cells.

Isolation and cloning of genes encoding glycerol dehydratase (*dhaB*) and 1,3-propanediol oxidoreductase (*dhaT*)

Cosmid vectors and cosmid transformation methods were used within the context of the present invention to clone large segments of genomic DNA from bacterial genera known to possess genes capable of processing glycerol to 1,3-propanediol. Specifically, genomic DNA from *K. pneumoniae* ATCC 25955 was isolated by methods well known in the art and digested with the restriction enzyme *Sau3A* for insertion into a cosmid vector Supercos 1 and packaged using GigapackII packaging extracts. Following construction of the vector *E. coli* XL1-Blue MR cells were transformed with the cosmid DNA. Transformants were screened for the ability to convert glycerol to 1,3-propanediol by growing the cells in the presence of glycerol and analyzing the media for 1,3-propanediol formation.

Two of the 1,3-propanediol positive transformants were analyzed and the cosmids were named pKP1 and pKP2. DNA sequencing revealed extensive homology to the glycerol dehydratase gene (*dhaB*) from *C. freundii*, demonstrating that these transformants contained DNA encoding the glycerol dehydratase gene. Other 1,3-propanediol positive transformants were analyzed and the cosmids were named pKP4 and pKP5. DNA sequencing revealed that these cosmids carried DNA encoding a diol dehydratase gene.

Isolation of genes encoding protein X, protein 1, protein 2 and protein 3

Although the instant invention utilizes the isolated genes from within a *Klebsiella* cosmid, alternate sources of dehydratase genes and protein X and protein 1, protein 2 and protein 3 include, but are not limited to, *Citrobacter*, *Clostridia*, and *Salmonella*. Tobimatsu, et al., 1996, J. Biol. Chem. 271: 22352-22357 disclose the *K. pneumoniae* glycerol dehydratase operon where protein X is identified as ORF 4; Segfried et al., 1996, J. Bacteriol. 178: 5793-5796 disclose the *C. freundii* glycerol dehydratase operon where protein X is identified as ORF Z. Figure 8 discloses

-- 18 --

*Klebsiella* diol dehydratase protein X and Figures 3, 4 and 5 disclose amino acid sequences of proteins 1, 2 and 3 from *Klebsiella* and *Citrobacter*.

Genes encoding G3PDH and G3P phosphatase

The present invention provides genes suitable for the expression of G3PDH and G3P  
5 phosphatase activities in a host cell.

Genes encoding G3PDH are known. For example, GPD1 has been isolated from *Saccharomyces* and has the base sequence given by SEQ ID NO:5, encoding the amino acid sequence given in SEQ ID NO:11 (Wang et al., *supra*). Similarly, G3PDH activity is has also been isolated from *Saccharomyces* encoded by GPD2 having the base sequence given in SEQ  
10 ID NO:6, encoding the amino acid sequence given in SEQ ID NO:12 (Eriksson et al., *Mol. Microbiol.* 17, 95, (1995).

It is contemplated that any gene encoding a polypeptide responsible for G3PDH activity is suitable for the purposes of the present invention wherein that activity is capable of catalyzing the conversion of dihydroxyacetone phosphate (DHAP) to glycerol-3-phosphate (G3P). Further, it is  
15 contemplated that any gene encoding the amino acid sequence of G3PDH as given by any one of SEQ ID NOS:11, 12, 13, 14, 15 and 16 corresponding to the genes GPD1, GPD2, GUT2, *gpsA*, *glpD*, and the  $\alpha$  subunit of *glpABC*, respectively, will be functional in the present invention wherein that amino acid sequence encompasses amino acid substitutions, deletions or additions that do not alter the function of the enzyme. It will be appreciated by the skilled person that genes  
20 encoding G3PDH isolated from other sources are also be suitable for use in the present invention. For example, genes isolated from prokaryotes include GenBank accessions M34393, M20938, L06231, U12567, L45246, L45323, L45324, L45325, U32164, and U39682; genes isolated from fungi include GenBank accessions U30625, U30876 and X56162; genes isolated from insects include GenBank accessions X61223 and X14179; and genes isolated from mammalian sources  
25 include GenBank accessions U12424, M25558 and X78593.

Genes encoding G3P phosphatase are known. For example, GPP2 has been isolated from *Saccharomyces cerevisiae* and has the base sequence given by SEQ ID NO:9 which encodes the amino acid sequence given in SEQ ID NO:17 (Norbeck et al., *J. Biol. Chem.* 271, p. 13875, 1996).

30 It is contemplated that any gene encoding a G3P phosphatase activity is suitable for the purposes of the present invention wherein that activity is capable of catalyzing the conversion of glycerol-3-phosphate to glycerol. Further, it is contemplated that any gene encoding the amino acid sequence of G3P phosphatase as given by SEQ ID NOS:33 and 17 will be functional in the present invention wherein that amino acid sequence encompasses amino acid substitutions,  
35 deletions or additions that do not alter the function of the enzyme. It will be appreciated by the skilled person that genes encoding G3P phosphatase isolated from other sources are also



-- 19 --

suitable for use in the present invention. For example, the dephosphorylation of glycerol-3-phosphate to yield glycerol may be achieved with one or more of the following general or specific phosphatases: alkaline phosphatase (EC 3.1.3.1) [GenBank M19159, M29663; U02550 or M33965]; acid phosphatase (EC 3.1.3.2) [GenBank U51210, U19789, U28658 or L20566];  
5 glycerol-3-phosphatase (EC 3.1.3.-) [GenBank Z38060 or U18813x11]; glucose-1-phosphatase (EC 3.1.3.10) [GenBank M33807]; glucose-6-phosphatase (EC 3.1.3.9) [GenBank U00445]; fructose-1,6-bisphosphatase (EC 3.1.3.11) [GenBank X12545 or J03207] or phosphatidyl glycerol phosphate phosphatase (EC 3.1.3.27) [GenBank M23546 and M23628].

Genes encoding glycerol kinase are known. For example, GUT1 encoding the glycerol  
10 kinase from *Saccharomyces* has been isolated and sequenced (Pavlik et al., *Curr. Genet.* 24, 21, (1993)) and the base sequence is given by SEQ ID NO:10 which encodes the amino acid sequence given in SEQ ID NO:18. It will be appreciated by the skilled artisan that although glycerol kinase catalyzes the degradation of glycerol in nature the same enzyme will be able to function in the synthesis of glycerol to convert glycerol-3-phosphate to glycerol under the  
15 appropriate reaction energy conditions. Evidence exists for glycerol production through a glycerol kinase. Under anaerobic or respiration-inhibited conditions, *Trypanosoma brucei* gives rise to glycerol in the presence of Glycerol-3-P and ADP. The reaction occurs in the glycosome compartment (D. Hammond, *J. Biol. Chem.* 260, 15646-15654, (1985)).

#### Host cells

20 Suitable host cells for the recombinant production of 1,3-propanediol may be either prokaryotic or eukaryotic and will be limited only by the host cell ability to express active enzymes. Preferred hosts will be those typically useful for production of glycerol or 1,3-propanediol such as *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*,  
25 *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*, *Bacillus*, *Streptomyces* and *Pseudomonas*. Most preferred in the present invention are *E. coli*, *Klebsiella* species and *Saccharomyces* species.

Adenosyl-cobalamin (coenzyme B<sub>12</sub>) is an essential cofactor for glycerol dehydratase activity. The coenzyme is the most complex non-polymeric natural product known, and its  
30 synthesis *in vivo* is directed using the products of about 30 genes. Synthesis of coenzyme B<sub>12</sub> is found in prokaryotes, some of which are able to synthesize the compound *de novo*, while others can perform partial reactions. *E. coli*, for example, cannot fabricate the corrin ring structure, but is able to catalyze the conversion of cobinamide to corrinoid and can introduce the 5'-deoxyadenosyl group.

35 Eukaryotes are unable to synthesize coenzyme B<sub>12</sub> *de novo* and instead transport vitamin B<sub>12</sub> from the extracellular milieu with subsequent conversion of the compound to its functional

-- 20 --

form of the compound by cellular enzymes. Three enzyme activities have been described for this series of reactions. 1) aquacobalamin reductase (EC 1.6.99.8) reduces Co(III) to Co(II); 2) cob(II)alamin reductase (EC 1.6.99.9) reduces Co(II) to Co(I); and 3) cob(I)alamin adenosyltransferase (EC 2.5.1.17) transfers a 5'deoxyadenosine moiety from ATP to the reduced

5 corrinoid. This last enzyme activity is the best characterized of the three, and is encoded by *cobA* in *S. typhimurium*, *btuR* in *E. coli* and *cobO* in *P. denitrificans*. These three cob(I)alamin adenosyltransferase genes have been cloned and sequenced. Cob(I)alamin adenosyltransferase activity has been detected in human fibroblasts and in isolated rat mitochondria (Fenton et al., *Biochem. Biophys. Res. Commun.* 98, 283-9, (1981)). The two enzymes involved in cobalt

10 reduction are poorly characterized and gene sequences are not available. There are reports of an aquacobalamin reductase from *Euglena gracilis* (Watanabe et al., *Arch. Biochem. Biophys.* 305, 421-7, (1993)) and a microsomal cob(III)alamin reductase is present in the microsomal and mitochondrial inner membrane fractions from rat fibroblasts (Pezacka, *Biochim. Biophys. Acta*, 1157, 167-77, (1993)).

15 Supplementing culture media with vitamin B<sub>12</sub> may satisfy the need to produce coenzyme B<sub>12</sub> for glycerol dehydratase activity in many microorganisms, but in some cases additional catalytic activities may have to be added or increased *in vivo*. Enhanced synthesis of coenzyme B<sub>12</sub> in eukaryotes may be particularly desirable. Given the published sequences for genes encoding cob(I)alamin adenosyltransferase, the cloning and expression of this gene could be

20 accomplished by one skilled in the art. For example, it is contemplated that yeast, such as *Saccharomyces*, could be constructed so as to contain genes encoding cob(I)alamin adenosyltransferase in addition to the genes necessary to effect conversion of a carbon substrate such as glucose to 1,3-propanediol. Cloning and expression of the genes for cobalt reduction requires a different approach. This could be based on a selection in *E. coli* for growth on

25 ethanolamine as sole N<sub>2</sub> source. In the presence of coenzyme B<sub>12</sub> ethanolamine ammonia-lyase enables growth of cells in the absence of other N<sub>2</sub> sources. If *E. coli* cells contain a cloned gene for cob(I)alamin adenosyltransferase and random cloned DNA from another organism, growth on ethanolamine in the presence of aquacobalamin should be enhanced and selected for if the random cloned DNA encodes cobalt reduction properties to facilitate adenosylation of

30 aquacobalamin.

Glycerol dehydratase is a multi-subunit enzyme consisting of three protein components which are arranged in an  $\alpha_2\beta_2\gamma_2$  configuration (M. Seyfried et al, *J. Bacteriol.*, 5793-5796 (1996)). This configuration is an inactive apo-enzyme which binds one molecule of coenzyme B<sub>12</sub> to become the catalytically active holo-enzyme. During catalysis, the holo-enzyme undergoes rapid,

35 first order inactivation, to become an inactive complex in which the coenzyme B<sub>12</sub> has been converted to hydroxycobalamin (Z. Schneider and J. Pawelkiewicz, *ACTA Biochim. Pol.* 311-328

-- 21 --

(1966)). Stoichiometric analysis of the reaction of glycerol dehydratase with glycerol as substrate revealed that each molecule of enzyme catalyzes 100,000 reactions before inactivation (Z. Schneider and J. Pawelkiewicz, ACTA Biochim. Pol. 311-328 (1966)). In vitro, this inactive complex can only be reactivated by removal of the hydroxycobalamin, by strong chemical treatment with magnesium and sulfite, and replacement with additional coenzyme B<sub>12</sub> (Z. Schneider et al., J. Biol. Chem. 3388-3396 (1970)). Inactivated glycerol dehydratase in wild type *Klebsiella pneumoniae* can be reactivated in situ (toluenized cells) in the presence of coenzyme B<sub>12</sub>, adenosine 5'-triphosphate (ATP), and manganese (S. Honda et al, J. Bacteriol. 1458-1465 (1980)). This reactivation was shown to be due to the ATP dependent replacement of the inactivated cobalamin with coenzyme B<sub>12</sub> (K. Ushio et al., J. Nutr. Sci. Vitaminol. 225-236 (1982)). Cell extract from toluenized cells which in situ catalyze the ATP, manganese, and coenzyme B<sub>12</sub> dependent reactivation are inactive with respect to this reactivation. Thus, without strong chemical reductive treatment or cell mediated replacement of the inactivated cofactor, glycerol dehydratase can only catalyzed 100,000 reactions per molecule.

The present invention demonstrates that the presence of protein X is important for in vivo reactivation of the dehydratase and the production of 1,3-propanediol is increased in a host cell capable of producing 1,3-propanediol in the presence of protein X. The present invention also discloses that the presence of protein 1, protein 2 and protein 3, in combination with protein X, also increased the production of 1,3-propanediol in a host cell capable of producing 1,3-propanediol.

In addition to *E. coli* and *Saccharomyces*, *Klebsiella* is a particularly preferred host. Strains of *Klebsiella pneumoniae* are known to produce 1,3-propanediol when grown on glycerol as the sole carbon. It is contemplated that *Klebsiella* can be genetically altered to produce 1,3-propanediol from monosaccharides, oligosaccharides, polysaccharides, or one-carbon substrates.

In order to engineer such strains, it will be advantageous to provide the *Klebsiella* host with the genes facilitating conversion of dihydroxyacetone phosphate to glycerol and conversion of glycerol to 1,3-propanediol either separately or together, under the transcriptional control of one or more constitutive or inducible promoters. The introduction of the DAR1 and GPP2 genes encoding glycerol-3-phosphate dehydrogenase and glycerol-3-phosphatase, respectively, will provide *Klebsiella* with genetic machinery to produce 1,3-propanediol from an appropriate carbon substrate.

The genes encoding protein X, protein 1, protein 2 and protein 3 or other enzymes associated with 1,3-propanediol production (e.g., G3PDH, G3P phosphatase, *dhaB* and/or *dhaT*) may be introduced on any plasmid vector capable of replication in *K. pneumoniae* or they may be integrated into the *K. pneumoniae* genome. For example, *K. pneumoniae* ATCC 25955 and

-- 22 --

*K. pneumoniae* ECL 2106 are known to be sensitive to tetracycline or chloramphenicol; thus plasmid vectors which are both capable of replicating in *K. pneumoniae* and encoding resistance to either or both of these antibiotics may be used to introduce these genes into *K. pneumoniae*. Methods of transforming *Klebsiella* with genes of interest are common and well known in the art and suitable protocols, including appropriate vectors and expression techniques may be found in Sambrook, *supra*.

#### Vectors and expression cassettes

The present invention provides a variety of vectors and transformation and expression cassettes suitable for the cloning, transformation and expression of protein X, protein 1, protein 2 and protein 3 as well as other proteins associated with 1,3-propanediol production, e.g., G3PDH and G3P phosphatase into a suitable host cell. Suitable vectors will be those which are compatible with the bacterium employed. Suitable vectors can be derived, for example, from a bacteria, a virus (such as bacteriophage T7 or a M-13 derived phage), a cosmid, a yeast or a plant. Protocols for obtaining and using such vectors are known to those in the art. (Sambrook et al., *Molecular Cloning: A Laboratory Manual* - volumes 1,2,3 (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, (1989)).

Typically, the vector or cassette contains sequences directing transcription and translation of the relevant gene, a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene which harbors transcriptional initiation controls and a region 3' of the DNA fragment which controls transcriptional termination. It is most preferred when both control regions are derived from genes homologous to the transformed host cell although it is to be understood that such control regions need not be derived from the genes native to the specific species chosen as a production host.

Initiation control regions or promoters, which are useful to drive expression of the protein x and protein 1, protein 2 or protein 3 in the desired host cell, are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving these genes is suitable for the present invention including but not limited to CYC1, HIS3, GAL1, GAL10, ADH1, PGK, PHO5, GAPDH, ADC1, TRP1, URA3, LEU2, ENO, TPI (useful for expression in *Saccharomyces*); AOX1 (useful for expression in *Pichia*); and lac, trp,  $IP_L$ ,  $IP_R$ , T7, tac, and trc (useful for expression in *E. coli*).

Termination control regions may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary, however, it is most preferred if included.

For effective expression of the instant enzymes, DNA encoding the enzymes are linked operably through initiation codons to selected expression control regions such that expression results in the formation of the appropriate messenger RNA.

-- 23 --

Transformation of suitable hosts and expression of genes for the  
production of 1,3-propanediol

Once suitable cassettes are constructed they are used to transform appropriate host cells. Introduction of the cassette containing dhaB activity, dhaB protein X and at least one of protein 1,  
5 protein 2 and protein 3 and optionally 1,3-propanediol oxidoreductase (*dhaT*), either separately or together, into the host cell may be accomplished by known procedures such as by transformation (e.g., using calcium-permeabilized cells, electroporation) or by transfection using a recombinant phage virus. (Sambrook et al., *supra.*). In the present invention, *E. coli* DH5a was transformed with dhaB subunits 1, 2 and 3 and dha protein X.

10 Additionally, *E. coli* W2042 (ATCC 98188) containing the genes encoding glycerol-3-phosphate dehydrogenase (G3PDH) and glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*) was created. Additionally, *S. cerevisiae* YPH500 (ATCC 74392) harboring plasmids pMCK10, pMCK17, pMCK30 and  
15 pMCK35 containing genes encoding glycerol-3-phosphate dehydrogenase (G3PDH) and glycerol-3-phosphatase (G3P phosphatase), glycerol dehydratase (*dhaB*), and 1,3-propanediol oxidoreductase (*dhaT*) was constructed. Both the above-mentioned transformed *E. coli* and *Saccharomyces* represent preferred embodiments of the invention.

Media and Carbon Substrates:

Fermentation media in the present invention must contain suitable carbon substrates.

20 Suitable substrates may include but are not limited to monosaccharides such as glucose and fructose, oligosaccharides such as lactose or sucrose, polysaccharides such as starch or cellulose, or mixtures thereof, and unpurified mixtures from renewable feedstocks such as cheese whey permeate, cornsteep liquor, sugar beet molasses, and barley malt. Additionally, the carbon  
25 substrate may also be one-carbon substrates such as carbon dioxide, or methanol for which metabolic conversion into key biochemical intermediates has been demonstrated. Glycerol production from single carbon sources (e.g., methanol, formaldehyde, or formate) has been reported in methylotrophic yeasts (Yamada et al., *Agric. Biol. Chem.*, 53(2) 541-543, (1989)) and in bacteria (Hunter et.al., *Biochemistry*, 24, 4148-4155, (1985)). These organisms can assimilate single carbon compounds, ranging in oxidation state from methane to formate, and produce  
30 glycerol. The pathway of carbon assimilation can be through ribulose monophosphate, through serine, or through xylulose-momophosphate (Gottschalk, Bacterial Metabolism, Second Edition, Springer-Verlag: New York (1986)). The ribulose monophosphate pathway involves the condensation of formate with ribulose-5-phosphate to form a 6 carbon sugar that becomes fructose and eventually the three carbon product glyceraldehyde-3-phosphate. Likewise, the  
35 serine pathway assimilates the one-carbon compound into the glycolytic pathway via methylenetetrahydrofolate.

-- 24 --

In addition to utilization of one and two carbon substrates, methylotrophic organisms are also known to utilize a number of other carbon-containing compounds such as methylamine, glucosamine and a variety of amino acids for metabolic activity. For example, methylotrophic yeast are known to utilize the carbon from methylamine to form trehalose or glycerol (Bellion et al., *Microb. Growth C1 Compd.*, [Int. Symp.], 7th (1993), 415-32. Editor(s): Murrell, J. Collin; Kelly, Don P. Publisher: Intercept, Andover, UK). Similarly, various species of *Candida* will metabolize alanine or oleic acid (Sulter et al., *Arch. Microbiol.*, 153(5), 485-9 (1990)). Hence, the source of carbon utilized in the present invention may encompass a wide variety of carbon-containing substrates and will only be limited by the requirements of the host organism.

Although it is contemplated that all of the above mentioned carbon substrates and mixtures thereof are suitable in the present invention, preferred carbon substrates are monosaccharides, oligosaccharides, polysaccharides, and one-carbon substrates. More preferred are sugars such as glucose, fructose, sucrose and single carbon substrates such as methanol and carbon dioxide. Most preferred is glucose.

In addition to an appropriate carbon source, fermentation media must contain suitable minerals, salts, cofactors, buffers and other components, known to those skilled in the art, suitable for the growth of the cultures and promotion of the enzymatic pathway necessary for glycerol production. Particular attention is given to Co(II) salts and/or vitamin B<sub>12</sub> or precursors thereof.

#### Culture Conditions:

Typically, cells are grown at 30 °C in appropriate media. Preferred growth media in the present invention are common commercially prepared media such as Luria Bertani (LB) broth, Sabouraud Dextrose (SD) broth or Yeast Malt Extract (YM) broth. Other defined or synthetic growth media may also be used and the appropriate medium for growth of the particular microorganism will be known by someone skilled in the art of microbiology or fermentation science. The use of agents known to modulate catabolite repression directly or indirectly, e.g., cyclic adenosine 2':3'-monophosphate or cyclic adenosine 2':5'-monophosphate, may also be incorporated into the reaction media. Similarly, the use of agents known to modulate enzymatic activities (e.g., sulphites, bisulphites and alkalis) that lead to enhancement of glycerol production may be used in conjunction with or as an alternative to genetic manipulations.

Suitable pH ranges for the fermentation are between pH 5.0 to pH 9.0, where pH 6.0 to pH 8.0 is preferred as range for the initial condition.

Reactions may be performed under aerobic or anaerobic conditions where anaerobic or microaerobic conditions are preferred.

-- 25 --

Batch and Continuous Fermentations:

The present process uses a batch method of fermentation. A classical batch fermentation is a closed system where the composition of the media is set at the beginning of the fermentation and not subject to artificial alterations during the fermentation. Thus, at the beginning of the fermentation the media is inoculated with the desired organism or organisms and fermentation is permitted to occur adding nothing to the system. Typically, however, a batch fermentation is "batch" with respect to the addition of the carbon source and attempts are often made at controlling factors such as pH and oxygen concentration. The metabolite and biomass compositions of the batch system change constantly up to the time the fermentation is stopped. Within batch cultures cells moderate through a static lag phase to a high growth log phase and finally to a stationary phase where growth rate is diminished or halted. If untreated, cells in the stationary phase will eventually die. Cells in log phase generally are responsible for the bulk of production of end product or intermediate.

A variation on the standard batch system is the Fed-Batch fermentation system which is also suitable in the present invention. In this variation of a typical batch system, the substrate is added in increments as the fermentation progresses. Fed-Batch systems are useful when catabolite repression is apt to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the media. Measurement of the actual substrate concentration in Fed-Batch systems is difficult and is therefore estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases such as CO<sub>2</sub>. Batch and Fed-Batch fermentations are common and well known in the art and examples may be found in Brock, *supra*.

It is also contemplated that the method would be adaptable to continuous fermentation methods. Continuous fermentation is an open system where a defined fermentation media is added continuously to a bioreactor and an equal amount of conditioned media is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth.

Continuous fermentation allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, one method will maintain a limiting nutrient such as the carbon source or nitrogen level at a fixed rate and allow all other parameters to moderate. In other systems a number of factors affecting growth can be altered continuously while the cell concentration, measured by media turbidity, is kept constant. Continuous systems strive to maintain steady state growth conditions and thus the cell loss due to media being drawn off must be balanced against the cell growth rate in the fermentation.

Methods of modulating nutrients and growth factors for continuous fermentation processes as

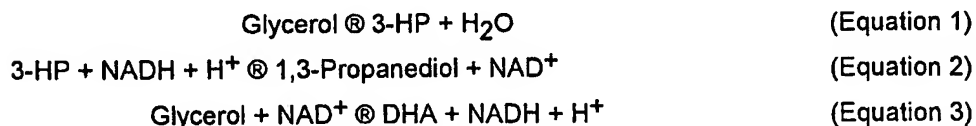
-- 26 --

well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology and a variety of methods are detailed by Brock, *supra*.

The present invention may be practiced using either batch, fed-batch or continuous processes and that any known mode of fermentation would be suitable. Additionally, it is contemplated that cells may be immobilized on a substrate as whole cell catalysts and subjected to fermentation conditions for 1,3-propanediol production.

Alterations in the 1,3-propanediol production pathway:

Representative enzyme pathway. The production of 1,3-propanediol from glucose can be accomplished by the following series of steps. This series is representative of a number of pathways known to those skilled in the art. Glucose is converted in a series of steps by enzymes of the glycolytic pathway to dihydroxyacetone phosphate (DHAP) and 3-phosphoglyceraldehyde (3-PG). Glycerol is then formed by either hydrolysis of DHAP to dihydroxyacetone (DHA) followed by reduction, or reduction of DHAP to glycerol 3-phosphate (G3P) followed by hydrolysis. The hydrolysis step can be catalyzed by any number of cellular phosphatases which are known to be specific or non-specific with respect to their substrates or the activity can be introduced into the host by recombination. The reduction step can be catalyzed by a  $\text{NAD}^+$  (or  $\text{NADP}^+$ ) linked host enzyme or the activity can be introduced into the host by recombination. It is notable that the *dha* regulon contains a glycerol dehydrogenase (E.C. 1.1.1.6) which catalyzes the reversible reaction of Equation 3.



Glycerol is converted to 1,3-propanediol via the intermediate 3-hydroxypropionaldehyde (3-HP) as has been described in detail above. The intermediate 3-HP is produced from glycerol (Equation 1) by a dehydratase enzyme which can be encoded by the host or can introduced into the host by recombination. This dehydratase can be glycerol dehydratase (E.C. 4.2.1.30), diol dehydratase (E.C. 4.2.1.28), or any other enzyme able to catalyze this transformation. Glycerol dehydratase, but not diol dehydratase, is encoded by the *dha* regulon. 1,3-Propanediol is produced from 3-HP (Equation 2) by a  $\text{NAD}^+$ - (or  $\text{NADP}^+$ ) linked host enzyme or the activity can introduced into the host by recombination. This final reaction in the production of 1,3-propanediol can be catalyzed by 1,3-propanediol dehydrogenase (E.C. 1.1.1.202) or other alcohol dehydrogenases.

Mutations and transformations that affect carbon channeling. A variety of mutant organisms comprising variations in the 1,3-propanediol production pathway will be useful in the present



-- 27 --

invention. The introduction of a triosephosphate isomerase mutation (*tpi*-) into the microorganism is an example of the use of a mutation to improve the performance by carbon channeling. Alternatively, mutations which diminish the production of ethanol (*adh*) or lactate (*ldh*) will increase the availability of NADH for the production of 1,3-propanediol. Additional mutations in  
5 steps of glycolysis after glyceraldehyde-3-phosphate such as phosphoglycerate mutase (*pgm*) would be useful to increase the flow of carbon to the 1,3-propanediol production pathway. Mutations that effect glucose transport such as PTS which would prevent loss of PEP may also prove useful. Mutations which block alternate pathways for intermediates of the 1,3-propanediol production pathway such as the glycerol catabolic pathway (*glp*) would also be useful to the  
10 present invention. The mutation can be directed toward a structural gene so as to impair or improve the activity of an enzymatic activity or can be directed toward a regulatory gene so as to modulate the expression level of an enzymatic activity.

Alternatively, transformations and mutations can be combined so as to control particular enzyme activities for the enhancement of 1,3-propanediol production. Thus it is within the scope  
15 of the present invention to anticipate modifications of a whole cell catalyst which lead to an increased production of 1,3-propanediol.

Identification and purification of 1,3-propanediol:

Methods for the purification of 1,3-propanediol from fermentation media are known in the art. For example, propanediols can be obtained from cell media by subjecting the reaction  
20 mixture to extraction with an organic solvent, distillation and column chromatography (U.S. 5,356,812). A particularly good organic solvent for this process is cyclohexane (U.S. 5,008,473).

1,3-Propanediol may be identified directly by submitting the media to high pressure liquid chromatography (HPLC) analysis. Preferred in the present invention is a method where  
25 fermentation media is analyzed on an analytical ion exchange column using a mobile phase of 0.01 N sulfuric acid in an isocratic fashion.

Identification and purification of G3PDH and G3P phosphatase:

The levels of expression of the proteins G3PDH and G3P phosphatase are measured by enzyme assays, G3PDH activity assay relied on the spectral properties of the cosubstrate, NADH,  
30 in the DHAP conversion to G-3-P. NADH has intrinsic UV/vis absorption and its consumption can be monitored spectrophotometrically at 340 nm. G3P phosphatase activity can be measured by any method of measuring the inorganic phosphate liberated in the reaction. The most commonly used detection method used the visible spectroscopic determination of a blue-colored phosphomolybdate ammonium complex.

-- 28 --

## EXAMPLES

### GENERAL METHODS

Procedures for phosphorylations, ligations and transformations are well known in the art. Techniques suitable for use in the following examples may be found in Sambrook, J. et al.,  
5 Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in Manual of Methods for General Bacteriology (Phillipp Gerhardt, R. G. E. Murray, Ralph N.  
10 Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds), American Society for Microbiology, Washington, DC. (1994)) or by Thomas D. Brock in Biotechnology: A Textbook of Industrial Microbiology, Second Edition, Sinauer Associates, Inc., Sunderland, MA (1989). All reagents and materials used for the growth and maintenance of bacterial cells were obtained from Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories  
15 (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), or Sigma Chemical Company (St. Louis, MO) unless otherwise specified.

The meaning of abbreviations is as follows: "h" means hour(s), "min" means minute(s), "sec" means second(s), "d" means day(s), "mL" means milliliters, "L" means liters.

### ENZYME ASSAYS

20 Glycerol dehydratase activity in cell-free extracts was determined using 1,2-propanediol as substrate. The assay, based on the reaction of aldehydes with methylbenzo-2-thiazolone hydrazone, has been described by Forage and Foster (*Biochim. Biophys. Acta*, 569, 249 (1979)). The activity of 1,3-propanediol oxidoreductase, sometimes referred to as 1,3-propanediol dehydrogenase, was determined in solution or in slab gels using 1,3-propanediol and NAD<sup>+</sup> as  
25 substrates as has also been described. Johnson and Lin, *J. Bacteriol.*, 169, 2050 (1987). NADH or NADPH dependent glycerol 3-phosphate dehydrogenase (G3PDH) activity was determined spectrophotometrically, following the disappearance of NADH or NADPH as has been described. (R. M. Bell and J. E. Cronan, Jr., *J. Biol. Chem.* 250:7153-8 (1975)).

Honda et al. (1980, In Situ Reactivation of Glycerol-Inactivated Coenzyme B<sub>12</sub>-Dependent  
30 Enzymes, Glycerol Dehydratase and Diol Dehydratase. *Journal of Bacteriology* 143:1458-1465) disclose an assay that measures the reactivation of dehydratases.

#### Assay for glycerol-3-phosphatase, GPP

The assay for enzyme activity was performed by incubating the extract with an organic phosphate substrate in a bis-Tris or MES and magnesium buffer, pH 6.5. The substrate used  
35 was l-a-glycerol phosphate; d,l-a-glycerol phosphate. The final concentrations of the reagents in the assay are: buffer (20 mM, bis-Tris or 50 mM MES); MgCl<sub>2</sub> (10 mM); and substrate (20 mM).

-- 29 --

If the total protein in the sample was low and no visible precipitation occurs with an acid quench, the sample was conveniently assayed in the cuvette. This method involved incubating an enzyme sample in a cuvette that contained 20 mM substrate (50 mM, 200 mM), 50 mM MES, 10 mM  $\text{MgCl}_2$ , pH 6.5 buffer. The final phosphatase assay volume was 0.5 mL. The enzyme-containing sample was added to the reaction mixture; the contents of the cuvette were mixed and then the cuvette was placed in a circulating water bath at  $T = 37^\circ\text{C}$  for 5 to 120 min -- depending on whether the phosphatase activity in the enzyme sample ranged from 2 to 0.02 U/mL. The enzymatic reaction was quenched by the addition of the acid molybdate reagent (0.4 mL). After the Fiske SubbaRow reagent (0.1 mL) and distilled water (1.5 mL) were added, the solution was mixed and allowed to develop. After 10 min, the absorbance of the samples was read at 660 nm using a Cary 219 UV/Vis spectrophotometer. The amount of inorganic phosphate released was compared to a standard curve that was prepared by using a stock inorganic phosphate solution (0.65 mM) and preparing 6 standards with final inorganic phosphate concentrations ranging from 0.026 to 0.130 mmol/mL.

#### 15 Isolation and Identification 1,3-propanediol

The conversion of glycerol to 1,3-propanediol was monitored by HPLC. Analyses were performed using standard techniques and materials available to one skilled in the art of chromatography. One suitable method utilized a Waters Maxima 820 HPLC system using UV (210 nm) and RI detection. Samples were injected onto a Shodex SH-1011 column (8 mm x 300 mm, purchased from Waters, Milford, MA) equipped with a Shodex SH-1011P precolumn (6 mm x 50 mm), temperature controlled at  $50^\circ\text{C}$ , using 0.01 N  $\text{H}_2\text{SO}_4$  as mobile phase at a flow rate of 0.5 mL/min. When quantitative analysis was desired, samples were prepared with a known amount of trimethylacetic acid as external standard. Typically, the retention times of glycerol (RI detection), 1,3-propanediol (RI detection), and trimethylacetic acid (UV and RI detection) were 20.67 min, 26.08 min, and 35.03 min, respectively.

Production of 1,3-propanediol was confirmed by GC/MS. Analyses were performed using standard techniques and materials available to one of skill in the art of GC/MS. One suitable method utilized a Hewlett Packard 5890 Series II gas chromatograph coupled to a Hewlett Packard 5971 Series mass selective detector (EI) and a HP-INNOWax column (30 m length, 0.25 mm i.d., 0.25 micron film thickness). The retention time and mass spectrum of 1,3-propanediol generated were compared to that of authentic 1,3-propanediol ( $m/e$ : 57, 58).

An alternative method for GC/MS involved derivatization of the sample. To 1.0 mL of sample (e.g., culture supernatant) was added 30  $\mu\text{L}$  of concentrated (70% v/v) perchloric acid. After mixing, the sample was frozen and lyophilized. A 1:1 mixture of bis(trimethylsilyl)trifluoroacetamide:pyridine (300  $\mu\text{L}$ ) was added to the lyophilized material, mixed vigorously and placed at  $65^\circ\text{C}$  for one h. The sample was clarified of insoluble material by

-- 30 --

centrifugation. The resulting liquid partitioned into two phases, the upper of which was used for analysis. The sample was chromatographed on a DB-5 column (48 m, 0.25 mm I.D., 0.25  $\mu$ m film thickness; from J&W Scientific) and the retention time and mass spectrum of the 1,3-propanediol derivative obtained from culture supernatants were compared to that obtained from authentic standards. The mass spectrum of TMS-derivatized 1,3-propanediol contains the characteristic ions of 205, 177, 130 and 115 AMU.

#### EXAMPLE 1

#### CLONING AND TRANSFORMATION OF *E. COLI*/HOST CELLS WITH COSMID DNA FOR THE EXPRESSION OF 1,3-PROPANEDIOL

##### 10 Media

Synthetic S12 medium was used in the screening of bacterial transformants for the ability to make 1,3-propanediol. S12 medium contains: 10 mM ammonium sulfate, 50 mM potassium phosphate buffer, pH 7.0, 2 mM  $MgCl_2$ , 0.7 mM  $CaCl_2$ , 50  $\mu$ M  $MnCl_2$ , 1  $\mu$ M  $FeCl_3$ , 1  $\mu$ M  $ZnCl_2$ , 1.7  $\mu$ M  $CuSO_4$ , 2.5  $\mu$ M  $CoCl_2$ , 2.4  $\mu$ M  $Na_2MoO_4$ , and 2  $\mu$ M thiamine hydrochloride.

15 Medium A used for growth and fermentation consisted of: 10 mM ammonium sulfate; 50 mM MOPS/KOH buffer, pH 7.5; 5 mM potassium phosphate buffer, pH 7.5; 2 mM  $MgCl_2$ ; 0.7 mM  $CaCl_2$ ; 50  $\mu$ M  $MnCl_2$ ; 1  $\mu$ M  $FeCl_3$ ; 1  $\mu$ M  $ZnCl_2$ ; 1.72  $\mu$ M  $CuSO_4$ ; 2.53  $\mu$ M  $CoCl_2$ ; 2.42  $\mu$ M  $Na_2MoO_4$ ; 2  $\mu$ M thiamine hydrochloride; 0.01% yeast extract; 0.01% casamino acids; 0.8  $\mu$ g/mL vitamin B<sub>12</sub>; and 50  $\mu$ g/mL amp. Medium A was supplemented with either 0.2% glycerol or 0.2% glycerol plus 0.2% D-glucose as required.

##### 20 Cells:

*Klebsiella pneumoniae* ECL2106 (Ruch et al., *J. Bacteriol.*, 124, 348 (1975)), also known in the literature as *K. aerogenes* or *Aerobacter aerogenes*, was obtained from E. C. C. Lin (Harvard Medical School, Cambridge, MA) and was maintained as a laboratory culture.

25 *Klebsiella pneumoniae* ATCC 25955 was purchased from American Type Culture Collection (Rockville, MD).

*E. coli* DH5a was purchased from Gibco/BRL and was transformed with the cosmid DNA isolated from *Klebsiella pneumoniae* ATCC 25955 containing a gene coding for either a glycerol or diol dehydratase enzyme. Cosmids containing the glycerol dehydratase were identified as pKP1 and pKP2 and cosmid containing the diol dehydratase enzyme were identified as pKP4. Transformed DH5a cells were identified as DH5a-pKP1, DH5a-pKP2, and DH5a-pKP4.

30 *E. coli* ECL707 (Sprenger et al., *J. Gen. Microbiol.*, 135, 1255 (1989)) was obtained from E. C. C. Lin (Harvard Medical School, Cambridge, MA) and was similarly transformed with cosmid DNA from *Klebsiella pneumoniae*. These transformants were identified as ECL707-pKP1 and ECL707-pKP2, containing the glycerol dehydratase gene and ECL707-pKP4 containing the diol dehydratase gene.

-- 31 --

*E. coli* AA200 containing a mutation in the *tpi* gene (Anderson et al., *J. Gen Microbiol.*, 62, 329 (1970)) was purchased from the *E. coli* Genetic Stock Center, Yale University (New Haven, CT) and was transformed with *Klebsiella* cosmid DNA to give the recombinant organisms AA200-pKP1 and AA200-pKP2, containing the glycerol dehydratase gene, and AA200-pKP4, containing the diol dehydratase gene.

#### DH5a:

Six transformation plates containing approximately 1,000 colonies of *E. coli* XL1-Blue MR transfected with *K. pneumoniae* DNA were washed with 5 mL LB medium and centrifuged. The bacteria were pelleted and resuspended in 5 mL LB medium + glycerol. An aliquot (50  $\mu$ L) was inoculated into a 15 mL tube containing S12 synthetic medium with 0.2% glycerol + 400 ng per mL of vitamin B<sub>12</sub> + 0.001% yeast extract + 50amp. The tube was filled with the medium to the top and wrapped with parafilm and incubated at 30 °C. A slight turbidity was observed after 48 h. Aliquots, analyzed for product distribution as described above at 78 h and 132 h, were positive for 1,3-propanediol, the later time points containing increased amounts of 1,3-propanediol.

The bacteria, testing positive for 1,3-propanediol production, were serially diluted and plated onto LB-50amp plates in order to isolate single colonies. Forty-eight single colonies were isolated and checked again for the production of 1,3-propanediol. Cosmid DNA was isolated from 6 independent clones and transformed into *E. coli* strain DH5a. The transformants were again checked for the production of 1,3-propanediol. Two transformants were characterized further and designated as DH5a-pKP1 and DH5a-pKP2.

A 12.1 kb EcoRI-SalI fragment from pKP1, subcloned into pIBI31 (IBI Biosystem, New Haven, CT), was sequenced and termed pHK28-26 (SEQ ID NO:19). Sequencing revealed the loci of the relevant open reading frames of the *dha* operon encoding glycerol dehydratase and genes necessary for regulation. Referring to SEQ ID NO:19, a fragment of the open reading frame for *dhaK* encoding dihydroxyacetone kinase is found at bases 1-399; the open reading frame *dhaD* encoding glycerol dehydrogenase is found at bases 983-2107; the open reading frame *dhaR* encoding the repressor is found at bases 2209-4134; the open reading frame *dhaT* encoding 1,3-propanediol oxidoreductase is found at bases 5017-6180; the open reading frame *dhaB1* encoding the alpha subunit glycerol dehydratase is found at bases 7044-8711; the open reading frame *dhaB2* encoding the beta subunit glycerol dehydratase is found at bases 8724-9308; the open reading frame *dhaB3* encoding the gamma subunit glycerol dehydratase is found at bases 9311-9736; and the open reading frame *dhaBX*, encoding a protein of unknown function is found at bases 9749-11572.

Single colonies of *E. coli* XL1-Blue MR transfected with packaged cosmid DNA from *K. pneumoniae* were inoculated into microtiter wells containing 200  $\mu$ L of S15 medium (ammonium sulfate, 10 mM; potassium phosphate buffer, pH 7.0, 1 mM; MOPS/KOH buffer,

-- 32 --

pH 7.0, 50 mM; MgCl<sub>2</sub>, 2 mM; CaCl<sub>2</sub>, 0.7 mM; MnCl<sub>2</sub>, 50 uM; FeCl<sub>3</sub>, 1 uM; ZnCl, 1 uM; CuSO<sub>4</sub>, 1.72 uM; CoCl<sub>2</sub>, 2.53 uM; Na<sub>2</sub>MoO<sub>4</sub>, 2.42 uM; and thiamine hydrochloride, 2 uM) + 0.2% glycerol + 400 ng/mL of vitamin B<sub>12</sub> + 0.001% yeast extract + 50 ug/mL ampicillin. In addition to the microtiter wells, a master plate containing LB-50 amp was also inoculated. After 96 h, 100 uL was withdrawn and centrifuged in a Rainin microfuge tube containing a 0.2 micron nylon membrane filter. Bacteria were retained and the filtrate was processed for HPLC analysis. Positive clones demonstrating 1,3-propanediol production were identified after screening approximately 240 colonies. Three positive clones were identified, two of which had grown on LB-50 amp and one of which had not. A single colony, isolated from one of the two positive clones grown on LB-50 amp and verified for the production of 1,3-propanediol, was designated as pKP4. Cosmid DNA was isolated from *E. coli* strains containing pKP4 and *E. coli* strain DH5a was transformed. An independent transformant, designated as DH5a-pKP4, was verified for the production of 1,3-propanediol.

#### ECL707:

*E. coli* strain ECL707 was transformed with cosmid *K. pneumoniae* DNA corresponding to one of pKP1, pKP2, pKP4 or the Supercos vector alone and named ECL707-pKP1, ECL707-pKP2, ECL707-pKP4, and ECL707-sc, respectively. ECL707 is defective in *glpK*, *gld*, and *ptsD* which encode the ATP-dependent glycerol kinase, NAD<sup>+</sup>-linked glycerol dehydrogenase, and enzyme II for dihydroxyacetone of the phosphoenolpyruvate-dependent phosphotransferase system, respectively.

Twenty single colonies of each cosmid transformation and five of the Supercos vector alone (negative control) transformation, isolated from LB-50amp plates, were transferred to a master LB-50amp plate. These isolates were also tested for their ability to convert glycerol to 1,3-propanediol in order to determine if they contained dehydratase activity. The transformants were transferred with a sterile toothpick to microtiter plates containing 200 uL of Medium A supplemented with either 0.2% glycerol or 0.2% glycerol plus 0.2% D-glucose. After incubation for 48 hr at 30 °C, the contents of the microtiter plate wells were filtered through an 0.45 micron nylon filter and chromatographed by HPLC. The results of these tests are given in Table 1.

Table 1

Conversion of glycerol to 1,3-propanediol by transformed ECL707

<u>Transformant</u>	<u>Glycerol*</u>	<u>Glycerol plus Glucose*</u>
ECL707-pKP1	19/20	19/20
ECL707-pKP2	18/20	20/20
ECL707-pKP4	0/20	20/20
ECL707-sc	0/5	0/5

-- 33 --

\*(Number of positive isolates/number of isolates tested)

#### AA200:

*E. coli* strain AA200 was transformed with cosmid *K. pneumoniae* DNA corresponding to one of pKP1, pKP2, pKP4 and the Supercos vector alone and named AA200-pKP1, AA200-pKP2, AA200-pKP4, and AA200-sc, respectively. Strain AA200 is defective in triosephosphate isomerase (*tpi*<sup>-</sup>).

Twenty single colonies of each cosmid transformation and five of the empty vector transformation were isolated and tested for their ability to convert glycerol to 1,3-propanediol as described for *E. coli* strain ECL707. The results of these tests are given in Table 2.

Table 2

Conversion of glycerol to 1,3-propanediol by transformed AA200

<u>Transformant</u>	<u>Glycerol*</u>	<u>Glycerol plus Glucose*</u>
AA200-pKP1	17/20	17/20
AA200-pKP2	17/20	17/20
AA200-pKP4	2/20	16/20
AA200-sc	0/5	0/5

\*(Number of positive isolates/number of isolates tested)

#### EXAMPLE 2

#### CONVERSION OF D-GLUCOSE TO 1,3-PROPANEDIOL BY RECOMBINANT *E. coli* USING

#### DAR1, GPP2, *dhaB*, and *dhaT*

#### Construction of general purpose expression plasmids for use in transformation of *Escherichia coli*

#### The expression vector pTacIQ

The *E. coli* expression vector, pTacIQ, contains the lacIq gene (Farabaugh, *Nature* 274, 5673 (1978)) and tac promoter (Amann et al., *Gene* 25, 167 (1983)) inserted into the EcoRI of pBR322 (Sutcliffe et al., *Cold Spring Harb. Symp. Quant. Biol.* 43, 77 (1979)). A multiple cloning site and terminator sequence (SEQ ID NO:20) replaces the pBR322 sequence from EcoRI to SphI.

#### Subcloning the glycerol dehydratase genes (*dhaB1*, 2, 3)

The open reading frame for *dhaB3* gene (incorporating an EcoRI site at the 5' end and a XbaI site at the 3' end) was amplified from pHK28-26 by PCR using primers (SEQ ID NOS:21 and 22). The product was subcloned into pLitmus29 (New England Biolab, Inc., Beverly, MA) to generate the plasmid pDHAB3 containing *dhaB3*.

-- 34 --

The region containing the entire coding region for the four genes of the *dhaB* operon from pHK28-26 was cloned into pBluescriptII KS+ (Stratagene, La Jolla, CA) using the restriction enzymes KpnI and EcoRI to create the plasmid pM7.

The *dhaBX* gene was removed by digesting the plasmid pM7, which contains  
5 *dhaB*(1,2,3,4), with Apal and XbaI (deleting part of *dhaB3* and all of *dhaBX*). The resulting 5.9 kb fragment was purified and ligated with the 325-bp Apal-XbaI fragment from plasmid pDHAB3 (restoring the *dhaB3* gene) to create pM11, which contains *dhaB*(1,2,3).

The open reading frame for the *dhaB1* gene (incorporating a HindIII site and a consensus RBS ribosome binding site at the 5' end and a XbaI site at the 3' end) was amplified from  
10 pHK28-26 by PCR using primers (SEQ ID NO:23 and SEQ ID NO:24). The product was subcloned into pLitmus28 (New England Biolab, Inc.) to generate the plasmid pDT1 containing *dhaB1*.

A NotI-XbaI fragment from pM11 containing part of the *dhaB1* gene, the *dhaB2* gene and the *dhaB3* gene was inserted into pDT1 to create the *dhaB* expression plasmid, pDT2. The  
15 HindIII-XbaI fragment containing the *dhaB*(1,2,3) genes from pDT2 was inserted into pTaclQ to create pDT3.

#### Subcloning the 1,3-propanediol dehydrogenase gene (*dhaT*)

The KpnI-SacI fragment of pHK28-26, containing the complete 1,3-propanediol dehydrogenase (*dhaT*) gene, was subcloned into pBluescriptII KS+ creating plasmid pAH1. The  
20 *dhaT* gene (incorporating an XbaI site at the 5' end and a BamHI site at the 3' end) was amplified by PCR from pAH1 as template DNA using synthetic primers (SEQ ID NO:25 with SEQ ID NO:26). The product was subcloned into pCR-Script (Stratagene) at the SrfI site to generate the plasmids pAH4 and pAH5 containing *dhaT*. The plasmid pAH4 contains the *dhaT* gene in the correct orientation for expression from the lac promoter in pCR-Script and pAH5 contains the  
25 *dhaT* gene in the opposite orientation. The XbaI-BamHI fragment from pAH4 containing the *dhaT* gene was inserted into pTaclQ to generate plasmid pAH8. The HindIII-BamHI fragment from pAH8 containing the RBS and *dhaT* gene was inserted into pBluescriptII KS+ to create pAH11. The HindIII-SalI fragment from pAH8 containing the RBS, *dhaT* gene and terminator was inserted into pBluescriptII SK+ to create pAH12.

#### Construction of an expression cassette for *dhaB*(1,2,3) and *dhaT*

An expression cassette for the *dhaB*(1,2,3) and *dhaT* was assembled from the individual *dhaB*(1,2,3) and *dhaT* subclones described above using standard molecular biology methods. The SpeI-KpnI fragment from pAH8 containing the RBS, *dhaT* gene and terminator was inserted into the XbaI-KpnI sites of pDT3 to create pAH23. The SmaI-EcoRI fragment between the *dhaB3*  
35 and *dhaT* gene of pAH23 was removed to create pAH26. The SpeI-NotI fragment containing an EcoRI site from pDT2 was used to replace the SpeI-NotI fragment of pAH26 to generate pAH27.



-- 35 --

Construction of expression cassette for *dhaT* and *dhaB*(1,2,3)

An expression cassette for *dhaT* and *dhaB*(1,2,3) was assembled from the individual *dhaB*(1,2,3) and *dhaT* subclones described previously using standard molecular biology methods. A *SpeI*-*SacI* fragment containing the *dhaB*(1,2,3) genes from pDT3 was inserted into pAH11 at the *SpeI*-*SacI* sites to create pAH24.

Cloning and expression of glycerol 3-phosphatase for increased glycerol production in *E. coli*

The *Saccharomyces cerevisiae* chromosome V lamda clone 6592 (Gene Bank, accession # U18813x11) was obtained from ATCC. The glycerol 3- phosphate phosphatase (GPP2) gene (incorporating an *Bam*HI-RBS-*Xba*I site at the 5' end and a *Sma*I site at the 3' end) was cloned by PCR cloning from the lamda clone as target DNA using synthetic primers (SEQ ID NO:27 with SEQ ID NO:28). The product was subcloned into pCR-Script (Stratagene) at the *Srf*I site to generate the plasmids pAH15 containing GPP2. The plasmid pAH15 contains the GPP2 gene in the inactive orientation for expression from the lac promoter in pCR-Script SK+. The *Bam*HI-*Sma*I fragment from pAH15 containing the GPP2 gene was inserted into pBlueScriptII SK+ to generate plasmid pAH19. The pAH19 contains the GPP2 gene in the correct orientation for expression from the lac promoter. The *Xba*I-*Pst*I fragment from pAH19 containing the GPP2 gene was inserted into pPHOX2 to create plasmid pAH21.

Plasmids for the expression of *dhaT*, *dhaB*(1,2,3) and GPP2 genes

A *Sal*I-*Eco*RI-*Xba*I linker (SEQ ID NOS:29 and 30) was inserted into pAH5 which was digested with the restriction enzymes, *Sal*I-*Xba*I to create pDT16. The linker destroys the *Xba*I site. The 1 kb *Sal*I-*Mlu*I fragment from pDT16 was then inserted into pAH24 replacing the existing *Sal*I-*Mlu*I fragment to create pDT18.

The 4.1 kb *Eco*RI-*Xba*I fragment containing the expression cassette for *dhaT* and *dhaB*(1,2,3) from pDT18 and the 1.0 kb *Xba*I-*Sal*I fragment containing the GPP2 gene from pAH21 was inserted into the vector pMMB66EH (Füste et al., *GENE*, 48, 119 (1986)) digested with the restriction enzymes *Eco*RI and *Sal*I to create pDT20.

Plasmids for the over-expression of DAR1 in *E. coli*

DAR1 was isolated by PCR cloning from genomic *S. cerevisiae* DNA using synthetic primers (SEQ ID NO:46 with SEQ ID NO:47). Successful PCR cloning places an *Nco*I site at the 5' end of DAR1 where the ATG within *Nco*I is the DAR1 initiator methionine. At the 3' end of DAR1 a *Bam*HI site is introduced following the translation terminator. The PCR fragments were digested with *Nco*I + *Bam*HI and cloned into the same sites within the expression plasmid pTrc99A (Pharmacia, Piscataway, New Jersey) to give pDAR1A.

In order to create a better ribosome binding site at the 5' end of DAR1, a *Spe*I-RBS-*Nco*I linker obtained by annealing synthetic primers (SEQ ID NO:48 with SEQ ID NO:49) was inserted into the *Nco*I site of pDAR1A to create pAH40. Plasmid pAH40 contains the new RBS and DAR1

-- 36 --

gene in the correct orientation for expression from the trc promoter of Trc99A (Pharmacia). The NcoI-BamHI fragment from pDAR1A and a second set of SpeI-RBS-NcoI linker obtained by annealing synthetic primers (SEQ ID NO:31 with SEQ ID NO:32) was inserted into the SpeI-BamHI site of pBluescript II-SK+ (Stratagene) to create pAH41. The construct pAH41 contains an ampicillin resistance gene. The NcoI-BamHI fragment from pDAR1A and a second set of SpeI-RBS-NcoI linker obtained by annealing synthetic primers (SEQ ID NO:31 with SEQ ID NO:32) was inserted into the SpeI-BamHI site of pBC-SK+ (Stratagene) to create pAH42. The construct pAH42 contains a chloramphenicol resistance gene.

#### Construction of an expression cassette for DAR1 and GPP2

An expression cassette for DAR1 and GPP2 was assembled from the individual DAR1 and GPP2 subclones described above using standard molecular biology methods. The BamHI-PstI fragment from pAH19 containing the RBS and GPP2 gene was inserted into pAH40 to create pAH43. The BamHI-PstI fragment from pAH19 containing the RBS and GPP2 gene was inserted into pAH41 to create pAH44. The same BamHI-PstI fragment from pAH19 containing the RBS and GPP2 gene was also inserted into pAH42 to create pAH45.

The ribosome binding site at the 5' end of GPP2 was modified as follows. A BamHI-RBS-SpeI linker, obtained by annealing synthetic primers GATCCAGGAAACAGA with CTAGTCTGTTTCCTG to the XbaI-PstI fragment from pAH19 containing the GPP2 gene, was inserted into the BamHI-PstI site of pAH40 to create pAH48. Plasmid pAH48 contains the DAR1 gene, the modified RBS, and the GPP2 gene in the correct orientation for expression from the trc promoter of pTrc99A (Pharmacia, Piscataway, N.J.).

#### E. coli strain construction

*E. coli* W1485 is a wild-type K-12 strain (ATCC 12435). This strain was transformed with the plasmids pDT20 and pAH42 and selected on LA (Luria Agar, Difco) plates supplemented with 50 mg/mL carbencillim and 10 mg/mL chloramphenicol.

#### Production of 1,3-propanediol from glucose

*E. coli* W1485/pDT20/pAH42 was transferred from a plate to 50 mL of a medium containing per liter: 22.5 g glucose, 6.85 g K<sub>2</sub>HPO<sub>4</sub>, 6.3 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.5 g NaHCO<sub>3</sub>, 2.5 g NaCl, 8 g yeast extract, 8 g tryptone, 2.5 mg vitamin B<sub>12</sub>, 2.5 mL modified Balch's trace-element solution, 50 mg carbencillim and 10 mg chloramphenicol, final pH 6.8 (HCl), then filter sterilized. The composition of modified Balch's trace-element solution can be found in Methods for General and Molecular Bacteriology (P. Gerhardt et al., eds, p. 158, American Society for Microbiology, Washington, DC (1994)). After incubating at 37 °C, 300 rpm for 6 h, 0.5 g glucose and IPTG (final concentration = 0.2 mM) were added and shaking was reduced to 100 rpm. Samples were analyzed by GC/MS. After 24 h, W1485/pDT20/pAH42 produced 1.1 g/L glycerol and 195 mg/L 1,3-propanediol.

-- 37 --

EXAMPLE 3CLONING AND EXPRESSION OF *dhaB* AND *dhaT*IN *Saccharomyces cerevisiae*

Expression plasmids that could exist as replicating episomal elements were constructed for each of the four *dha* genes. For all expression plasmids a yeast ADH1 promoter was present and separated from a yeast ADH1 transcription terminator by fragments of DNA containing recognition sites for one or more restriction endonucleases. Each expression plasmid also contained the gene for b-lactamase for selection in *E. coli* on media containing ampicillin, an origin of replication for plasmid maintenance in *E. coli*, and a 2 micron origin of replication for maintenance in *S. cerevisiae*. The selectable nutritional markers used for yeast and present on the expression plasmids were one of the following: HIS3 gene encoding imidazoleglycerolphosphate dehydratase, URA3 gene encoding orotidine 5'-phosphate decarboxylase, TRP1 gene encoding N-(5'-phosphoribosyl)-anthranilate isomerase, and LEU2 encoding b-isopropylmalate dehydrogenase.

The open reading frames for *dhaT*, *dhaB3*, *dhaB2* and *dhaB1* were amplified from pHK28-26 (SEQ ID NO:19) by PCR using primers (SEQ ID NO:38 with SEQ ID NO:39, SEQ ID NO:40 with SEQ ID NO:41, SEQ ID NO:42 with SEQ ID NO:43, and SEQ ID NO:44 with SEQ ID NO:45 for *dhaT*, *dhaB3*, *dhaB2* and *dhaB1*, respectively) incorporating EcoR1 sites at the 5' ends (10 mM Tris pH 8.3, 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, 0.0001% gelatin, 200 mM dATP, 200 mM dCTP, 200 mM dGTP, 200 mM dTTP, 1 mM each primer, 1-10 ng target DNA, 25 units/mL Amplitaq<sup>®</sup> DNA polymerase (Perkin-Elmer Cetus, Norwalk CT)). PCR parameters were 1 min at 94 °C, 1 min at 55 °C, 1 min at 72 °C, 35 cycles. The products were subcloned into the EcoR1 site of pHIL-D4 (Phillips Petroleum, Bartlesville, OK) to generate the plasmids pMP13, pMP14, pMP20 and pMP15 containing *dhaT*, *dhaB3*, *dhaB2* and *dhaB1*, respectively.

Construction of *dhaB1* expression plasmid pMCK10

The 7.8 kb replicating plasmid pGADGH (Clontech, Palo Alto, CA) was digested with HindIII, dephosphorylated, and ligated to the *dhaB1* HindIII fragment from pMP15. The resulting plasmid (pMCK10) had *dhaB1* correctly oriented for transcription from the ADH1 promoter and contained a LEU2 marker.

Construction of *dhaB2* expression plasmid pMCK17

Plasmid pGADGH (Clontech, Palo Alto, CA) was digested with HindIII and the single-strand ends converted to EcoRI ends by ligation with HindIII-XmnI and EcoRI-XmnI adaptors (New England Biolabs, Beverly, MA). Selection for plasmids with correct EcoRI ends was achieved by ligation to a kanamycin resistance gene on an EcoRI fragment from plasmid pUC4K (Pharmacia Biotech, Uppsala), transformation into *E. coli* strain DH5a and selection on LB plates containing 25 mg/mL kanamycin. The resulting plasmid (pGAD/KAN2) was digested with SnaBI

-- 38 --

and EcoRI and a 1.8 kb fragment with the ADH1 promoter was isolated. Plasmid pGBT9 (Clontech, Palo Alto, CA) was digested with SnaBI and EcoRI, and the 1.5 kb ADH1/GAL4 fragment replaced by the 1.8 kb ADH1 promoter fragment isolated from pGAD/KAN2 by digestion with SnaBI and EcoRI. The resulting vector (pMCK11) is a replicating plasmid in yeast with an ADH1 promoter and terminator and a TRP1 marker. Plasmid pMCK11 was digested with EcoRI, dephosphorylated, and ligated to the *dhaB2* EcoRI fragment from pMP20. The resulting plasmid (pMCK17) had *dhaB2* correctly oriented for transcription from the ADH1 promoter and contained a TRP1 marker.

#### Construction of *dhaB3* expression plasmid pMCK30

Plasmid pGBT9 (Clontech) was digested with NaeI and PvuII and the 1 kb TRP1 gene removed from this vector. The TRP1 gene was replaced by a URA3 gene donated as a 1.7 kb AatII/NaeI fragment from plasmid pRS406 (Stratagene) to give the intermediary vector pMCK32. The truncated ADH1 promoter present on pMCK32 was removed on a 1.5 kb SnaBI/EcoRI fragment, and replaced with a full-length ADH1 promoter on a 1.8 kb SnaBI/EcoRI fragment from plasmid pGAD/KAN2 to yield the vector pMCK26. The unique EcoRI site on pMCK26 was used to insert an EcoRI fragment with *dhaB3* from plasmid pMP14 to yield pMCK30. The pMCK30 replicating expression plasmid has *dhaB3* orientated for expression from the ADH1 promoter, and has a URA3 marker.

#### Construction of *dhaT* expression plasmid pMCK35

Plasmid pGBT9 (Clontech) was digested with NaeI and PvuII and the 1 kb TRP1 gene removed from this vector. The TRP1 gene was replaced by a HIS3 gene donated as an XmnI/NaeI fragment from plasmid pRS403 (Stratagene) to give the intermediary vector pMCK33. The truncated ADH1 promoter present on pMCK33 was removed on a 1.5 kb SnaBI/EcoRI fragment, and replaced with a full-length ADH1 promoter on a 1.8 kb SnaBI/EcoRI fragment from plasmid pGAD/KAN2 to yield the vector pMCK31. The unique EcoRI site on pMCK31 was used to insert an EcoRI fragment with *dhaT* from plasmid pMP13 to yield pMCK35. The pMCK35 replicating expression plasmid has *dhaT* orientated for expression from the ADH1 promoter, and has a HIS3 marker.

#### Transformation of *S. cerevisiae* with *dha* expression plasmids

*S. cerevisiae* strain YPH500 (*ura3-52 lys2-801 ade2-101 trp1-D63 his3-D200 leu2-D1*) (Sikorski R. S. and Hieter P., *Genetics* 122, 19-27, (1989)) purchased from Stratagene (La Jolla, CA) was transformed with 1-2 mg of plasmid DNA using a Frozen-EZ Yeast Transformation Kit (Catalog #T2001) (Zymo Research, Orange, CA). Colonies were grown on Supplemented Minimal Medium (SMM - 0.67% yeast nitrogen base without amino acids, 2% glucose) for 3-4 d at 29 °C with one or more of the following additions: adenine sulfate (20 mg/L), uracil (20 mg/L),

-- 39 --

L-tryptophan (20 mg/L), L-histidine (20 mg/L), L-leucine (30 mg/L), L-lysine (30 mg/L). Colonies were streaked on selective plates and used to inoculate liquid media.

#### Screening of *S. cerevisiae* transformants for *dha* genes

Chromosomal DNA from URA<sup>+</sup>, HIS<sup>+</sup>, TRP<sup>+</sup>, LEU<sup>+</sup> transformants was analyzed by PCR using primers specific for each gene (SEQ ID NOS:38-45). The presence of all four open reading frames was confirmed.

#### Expression of *dhaB* and *dhaT* activity in transformed *S. cerevisiae*

The presence of active glycerol dehydratase (*dhaB*) and 1,3-propanediol oxido-reductase (*dhaT*) was demonstrated using *in vitro* enzyme assays. Additionally, western blot analysis confirmed protein expression from all four open reading frames.

Strain YPH500, transformed with the group of plasmids pMCK10, pMCK17, pMCK30 and pMCK35, was grown on Supplemented Minimal Medium containing 0.67% yeast nitrogen base without amino acids 2% glucose 20 mg/L adenine sulfate, and 30 mg/L L-lysine. Cells were homogenized and extracts assayed for *dhaB* activity. A specific activity of 0.12 units per mg protein was obtained for glycerol dehydratase, and 0.024 units per mg protein for 1,3-propanediol oxido-reductase.

### EXAMPLE 4

#### PRODUCTION OF 1,3-PROPANEDIOL FROM D-GLUCOSE

##### USING RECOMBINANT *Saccharomyces cerevisiae*

*S. cerevisiae* YPH500, harboring the groups of plasmids pMCK10, pMCK17, pMCK30 and pMCK35, was grown in a BiostatB fermenter (B Braun Biotech, Inc.) in 1.0 L of minimal medium initially containing 20 g/L glucose, 6.7 g/L yeast nitrogen base without amino acids, 40 mg/L adenine sulfate and 60 mg/L L-lysine·HCl. During the course of the growth, an additional equivalent of yeast nitrogen base, adenine and lysine was added. The fermenter was controlled at pH 5.5 with addition of 10% phosphoric acid and 2 M NaOH, 30 °C, and 40% dissolved oxygen tension through agitation control. After 38 h, the cells (OD<sub>600</sub> = 5.8 AU) were harvested by centrifugation and resuspended in base medium (6.7 g/L yeast nitrogen base without amino acids, 20 mg/L adenine sulfate, 30 mg/L L-lysine·HCl, and 50 mM potassium phosphate buffer, pH 7.0).

Reaction mixtures containing cells (OD<sub>600</sub> = 20 AU) in a total volume of 4 mL of base media supplemented with 0.5% glucose, 5 ug/mL coenzyme B<sub>12</sub> and 0, 10, 20, or 40 mM chloroquine were prepared, in the absence of light and oxygen (nitrogen sparging), in 10 mL crimp sealed serum bottles and incubated at 30 °C with shaking. After 30 h, aliquots were withdrawn and analyzed by HPLC. The results are shown in the Table 3.

-- 40 --

Table 3Production of 1,3-propanediol using recombinant *S. cerevisiae*

reaction	chloroquine (mM)	1,3-propanediol (mM)
1	0	0.2
2	10	0.2
3	20	0.3
4	40	0.7

EXAMPLE 5

USE OF A *S. cerevisiae* DOUBLE TRANSFORMANT FOR PRODUCTION  
OF 1,3-PROPANEDIOL FROM D-GLUCOSE WHERE *dhaB* AND *dhaT* ARE  
INTEGRATED INTO THE GENOME

Example 5 prophetically demonstrates the transformation of *S. cerevisiae* with *dhaB1*, *dhaB2*, *dhaB3*, and *dhaT* and the stable integration of the genes into the yeast genome for the production of 1,3-propanediol from glucose.

Construction of expression cassettes

Four expression cassettes (*dhaB1*, *dhaB2*, *dhaB3*, and *dhaT*) are constructed for glucose-induced and high-level constitutive expression of these genes in yeast, *Saccharomyces cerevisiae*. These cassettes consist of: (i) the phosphoglycerate kinase (PGK) promoter from *S. cerevisiae* strain S288C; (ii) one of the genes *dhaB1*, *dhaB2*, *dhaB3*, or *dhaT*; and (iii) the PGK terminator from *S. cerevisiae* strain S288C. The PCR-based technique of gene splicing by overlap extension (Horton et al., *BioTechniques*, 8:528-535, (1990)) is used to recombine DNA sequences to generate these cassettes with seamless joints for optimal expression of each gene. These cassettes are cloned individually into a suitable vector (pLITMUS 39) with restriction sites amenable to multi-cassette cloning in yeast expression plasmids.

Construction of yeast integration vectors

Vectors used to effect the integration of expression cassettes into the yeast genome are constructed. These vectors contain the following elements: (i) a polycloning region into which expression cassettes are subcloned; (ii) a unique marker used to select for stable yeast transformants; (iii) replication origin and selectable marker allowing gene manipulation in *E. coli* prior to transforming yeast. One integration vector contains the *URA3* auxotrophic marker (Ylp352b), and a second integration vector contains the *LYS2* auxotrophic marker (pKP7).

-- 41 --

Construction of yeast expression plasmids

Expression cassettes for *dhaB1* and *dhaB2* are subcloned into the polycloning region of the Ylp352b (expression plasmid #1), and expression cassettes for *dhaB3* and *dhaT* are subcloned into the polycloning region of pKP7 (expression plasmid #2).

5 Transformation of yeast with expression plasmids

*S. cerevisiae* (*ura3*, *lys2*) is transformed with expression plasmid #1 using Frozen-EZ Yeast Transformation kit (Zymo Research, Orange, CA), and transformants selected on plates lacking uracil. Integration of expression cassettes for *dhaB1* and *dhaB2* is confirmed by PCR analysis of chromosomal DNA. Selected transformants are re-transformed with expression  
10 plasmid #2 using Frozen-EZ Yeast Transformation kit, and double transformants selected on plates lacking lysine. Integration of expression cassettes for *dhaB3* and *dhaT* is confirmed by PCR analysis of chromosomal DNA. The presence of all four expression cassettes (*dhaB1*, *dhaB2*, *dhaB3*, *dhaT*) in double transformants is confirmed by PCR analysis of chromosomal DNA.

15 Protein production from double-transformed yeast

Production of proteins encoded by *dhaB1*, *dhaB2*, *dhaB3* and *dhaT* from double-transformed yeast is confirmed by Western blot analysis.

Enzyme activity from double-transformed yeast

Active glycerol dehydratase and active 1,3-propanediol dehydrogenase from double-  
20 transformed yeast is confirmed by enzyme assay as described in General Methods above.

Production of 1,3-propanediol from double-transformed yeast

Production of 1,3-propanediol from glucose in double-transformed yeast is demonstrated essentially as described in Example 4.

EXAMPLE 625 CONSTRUCTION OF PLASMIDS CONTAINING DAR1/GPP2OR *dhaT/dhaB1-3* AND TRANSFORMATION INTO *KLEBSIELLA* SPECIES

*K. pneumoniae* (ATCC 25955), *K. pneumoniae* (ECL2106), and *K. oxytoca* (ATCC 8724) are naturally resistant to ampicillin (up to 150 ug/mL) and kanamycin (up to 50 ug/mL), but sensitive to tetracycline (10 ug/mL) and chloramphenicol (25 ug/mL). Consequently, replicating  
30 plasmids which encode resistance to these latter two antibiotics are potentially useful as cloning vectors for these *Klebsiella* strains. The wild-type *K. pneumoniae* (ATCC 25955), the glucose-derepressed *K. pneumonia* (ECL2106), and *K. oxytoca* (ATCC 8724) were successfully transformed to tetracycline resistance by electroporation with the moderate-copy-number plasmid, pBR322 (New England Biolabs, Beverly, MA). This was accomplished by the following  
35 procedure: Ten mL of an overnight culture was inoculated into 1 L LB (1% (w/v) Bacto-tryptone (Difco, Detroit, MI), 0.5% (w/v) Bacto-yeast extract (Difco) and 0.5% (w/v) NaCl (Sigma, St. Louis,

-- 42 --

MO) and the culture was incubated at 37 °C to an OD<sub>600</sub> of 0.5-0.7. The cells were chilled on ice, harvested by centrifugation at 4000 x g for 15 min, and resuspended in 1 L ice-cold sterile 10% glycerol. The cells were repeatedly harvested by centrifugation and progressively resuspended in 500 mL, 20 mL and, finally, 2 mL ice-cold sterile 10% glycerol. For  
5 electroporation, 40 uL of cells were mixed with 1-2 uL DNA in a chilled 0.2 cm cuvette and were pulsed at 200 Ω, 2.5 kV for 4-5 msec using a BioRad Gene Pulser (BioRad, Richmond, CA). One mL of SOC medium (2% (w/v) Bacto-tryptone (Difco), 0.5% (w/v) Bacto-yeast extract (Difco), 10 mM NaCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub>, 2.5 mM KCl and 20 mM glucose) was added to the cells and, after the suspension was transferred to a 17 x 100 mm sterile polypropylene tube, the  
10 culture was incubated for 1 hr at 37 °C, 225 rpm. Aliquots were plated on selective medium, as indicated. Analyses of the plasmid DNA from independent tetracycline-resistant transformants showed the restriction endonuclease digestion patterns typical of pBR322, indicating that the vector was stably maintained after overnight culture at 37 °C in LB containing tetracycline (10 ug/mL). Thus, this vector, and derivatives such as pBR329 (ATCC 37264) which encodes  
15 resistance to ampicillin, tetracycline and chloramphenicol, may be used to introduce the *DAR1/GPP2* and *dhaT/dhaB1-3* expression cassettes into *K. pneumoniae* and *K. oxytoca*.

The *DAR1* and *GPP2* genes may be obtained by PCR-mediated amplification from the *Saccharomyces cerevisiae* genome, based on their known DNA sequence. The genes are then transformed into *K. pneumoniae* or *K. oxytoca* under the control of one or more promoters that  
20 may be used to direct their expression in media containing glucose. For convenience, the genes were obtained on a 2.4 kb DNA fragment obtained by digestion of plasmid pAH44 with the *PvuII* restriction endonuclease, whereby the genes are already arranged in an expression cassette under the control of the *E. coli lac* promoter. This DNA fragment was ligated to *PvuII*-digested pBR329, producing the insertional inactivation of its chloramphenicol resistance gene. The  
25 ligated DNA was used to transform *E. coli* DH5α (Gibco, Gaithersburg, MD). Transformants were selected by their resistance to tetracycline (10 ug/mL) and were screened for their sensitivity to chloramphenicol (25 ug/mL). Analysis of the plasmid DNA from tetracycline-resistant, chloramphenicol-sensitive transformants confirmed the presence of the expected plasmids, in which the P<sub>lac</sub>-*dar1-gpp2* expression cassette was subcloned in either orientation into the  
30 pBR329 *PvuII* site. These plasmids, designated pJSP1A (clockwise orientation) and pJSP1B (counterclockwise orientation), were separately transformed by electroporation into *K. pneumonia* (ATCC 25955), *K. pneumonia* (ECL2106) and *K. oxytoca* (ATCC 8724) as described. Transformants were selected by their resistance to tetracycline (10 ug/mL) and were screened for their sensitivity to chloramphenicol (25 ug/mL). Restriction analysis of the plasmids isolated from  
35 independent transformants showed only the expected digestion patterns, and confirmed that they were stably maintained at 37 °C with antibiotic selection. The expression of the *DAR1* and *GPP2* genes may be enhanced by the addition of IPTG (0.2-2.0 mM) to the growth medium.



The four *K. pneumoniae* *dhaB*(1-3) and *dhaT* genes may be obtained by PCR-mediated amplification from the *K. pneumoniae* genome, based on their known DNA sequence. These genes are then transformed into *K. pneumoniae* under the control of one or more promoters that may be used to direct their expression in media containing glucose. For convenience, the genes  
 5 were obtained on an approximately 4.0 kb DNA fragment obtained by digestion of plasmid pAH24 with the *KpnI*/*SacI* restriction endonucleases, whereby the genes are already arranged in an expression cassette under the control of the *E. coli lac* promoter. This DNA fragment was ligated to similarly digested pBC-KS+ (Stratagene, LaJolla, CA) and used to transform *E. coli* DH5 $\alpha$ . Transformants were selected by their resistance to chloramphenicol (25 ug/mL) and were  
 10 screened for a white colony phenotype on LB agar containing X-gal. Restriction analysis of the plasmid DNA from chloramphenicol-resistant transformants demonstrating the white colony phenotype confirmed the presence of the expected plasmid, designated pJSP2, in which the *dhaT-dhaB*(1-3) genes were subcloned under the control of the *E. coli lac* promoter.

To enhance the conversion of glucose to 1,3-propanediol, this plasmid was separately  
 15 transformed by electroporation into *K. pneumoniae* (ATCC 25955) (pJSP1A), *K. pneumoniae* (ECL2106) (pJSP1A) and *K. oxytoca* (ATCC 8724) (pJSP1A) already containing the *P<sub>lac-dar1-gpp2</sub>* expression cassette. Cotransformants were selected by their resistance to both tetracycline (10 ug/mL) and chloramphenicol (25 ug/mL). Restriction analysis of the plasmids isolated from independent cotransformants showed the digestion patterns expected for both  
 20 pJSP1A and pJSP2. The expression of the *DAR1*, *GPP2*, *dhaB*(1-3), and *dhaT* genes may be enhanced by the addition of IPTG (0.2-2.0 mM) to the medium.

#### EXAMPLE 7

##### Production of 1,3 propanediol from glucose by *K. pneumoniae*

*Klebsiella pneumoniae* strains ECL 2106 and 2106-47, both transformed with pJSP1A,  
 25 and ATCC 25955, transformed with pJSP1A and pJSP2, were grown in a 5 L Applikon fermenter under various conditions (see Table 4) for the production of 1,3-propanediol from glucose. Strain 2104-47 is a fluoroacetate-tolerant derivative of ECL 2106 which was obtained from a fluoroacetate/lactate selection plate as described in Bauer et al., *Appl. Environ. Microbiol.* 56, 1296 (1990). In each case, the medium used contained 50-100 mM potassium phosphate buffer,  
 30 pH 7.5, 40 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.1% (w/v) yeast extract, 10  $\mu$ M CoCl<sub>2</sub>, 6.5  $\mu$ M CuCl<sub>2</sub>, 100  $\mu$ M FeCl<sub>3</sub>, 18  $\mu$ M FeSO<sub>4</sub>, 5  $\mu$ M H<sub>3</sub>BO<sub>3</sub>, 50  $\mu$ M MnCl<sub>2</sub>, 0.1  $\mu$ M Na<sub>2</sub>MoO<sub>4</sub>, 25  $\mu$ M ZnCl<sub>2</sub>, 0.82 mM MgSO<sub>4</sub>, 0.9 mM CaCl<sub>2</sub>, and 10-20 g/L glucose. Additional glucose was fed, with residual glucose maintained in excess. Temperature was controlled at 37 °C and pH controlled at 7.5 with 5N KOH or NaOH. Appropriate antibiotics were included for plasmid maintenance; IPTG  
 35 (isopropyl-b-D-thiogalactopyranoside) was added at the indicated concentrations as well. For anaerobic fermentations, 0.1 vvm nitrogen was sparged through the reactor; when the dO

-- 44 --

setpoint was 5%, 1 vvm air was sparged through the reactor and the medium was supplemented with vitamin B12. Final concentrations and overall yields (g/g) are shown in Table 4.

Table 4

Production of 1,3 propanediol from glucose by *K. pneumoniae*

Organism	dO	IPTG, mM	vitamin B12, mg/L	Titer, g/L	Yield, g/g
25955[pJSP1A/pJS P2]	0	0.5	0	8.1	16%
25955[pJSP1A/pJS P2]	5%	0.2	0.5	5.2	4%
2106[pJSP1A]	0	0	0	4.9	17%
2106[pJSP1A]	5%	0	5	6.5	12%
2106-47[pJSP1A]	5%	0.2	0.5	10.9	12%

5

## EXAMPLE 8

Conversion of carbon substrates to 1,3-propanediol by recombinant*K. pneumoniae* containing *dar1*, *gpp2*, *dhaB*, and *dhaT*

A. Conversion of D-fructose to 1,3-propanediol by various *K. pneumoniae* recombinant strains:

- 10 Single colonies of *K. pneumoniae* (ATCC 25955 pJSP1A), *K. pneumoniae* (ATCC 25955 pJSP1A/pJSP2), *K. pneumoniae* (ATCC 2106 pJSP1A), and *K. pneumoniae* (ATCC 2106 pJSP1A/pJSP2) were transferred from agar plates and in separate culture tubes were subcultured overnight in Luria-Bertani (LB) broth containing the appropriate antibiotic agent(s). A 50-mL flask containing 45 mL of a steri-filtered minimal medium defined as LLMM/F which
- 15 contains per liter: 10 g fructose; 1 g yeast extract; 50 mmoles potassium phosphate, pH 7.5; 40 mmoles (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; 0.09 mmoles calcium chloride; 2.38 mg CoCl<sub>2</sub>·6H<sub>2</sub>O; 0.88 mg CuCl<sub>2</sub>·2H<sub>2</sub>O; 27 mg FeCl<sub>3</sub>·6H<sub>2</sub>O; 5 mg FeSO<sub>4</sub>·7H<sub>2</sub>O; 0.31 mg H<sub>3</sub>BO<sub>3</sub>; 10 mg MnCl<sub>2</sub>·4H<sub>2</sub>O; 0.023 mg Na<sub>2</sub>MoO<sub>4</sub>·2H<sub>2</sub>O; 3.4 mg ZnCl<sub>2</sub>; 0.2 g MgSO<sub>4</sub>·7H<sub>2</sub>O. Tetracycline at 10 ug/mL was added to medium for reactions using either of the single plasmid recombinants; 10 ug/mL
- 20 tetracycline and 25 ug/mL chloramphenicol for reactions using either of the double plasmid recombinants. The medium was thoroughly sparged with nitrogen prior to inoculation with 2 mL of the subculture. IPTG (I) at final concentration of 0.5 mM was added to some flasks. The flasks were capped, then incubated at 37 °C, 100 rpm in a New Brunswick Series 25 incubator/shaker. Reactions were run for at least 24 hours or until most of the carbon substrate was converted into
- 25 products. Samples were analyzed by HPLC. Table 5 describes the yields of 1,3-propanediol (3G) produced from fructose by the various *Klebsiella* recombinants.

-- 45 --

Table 5

Production of 1,3-propanediol from D-fructose using recombinant *Klebsiella*

Klebsiella Strain	Medium	Conversion	[3G] (g/L)	Yield Carbon (%)
2106 pBR329	LLMM/F	100	0	0
2106 pJSP1A	LLMM/F	50	0.66	15.5
2106 pJSP1A	LLMM/F + I	100	0.11	1.4
2106 pJSP1A/pJSP2	LLMM/F	58	0.26	5
25955 pBR329	LLMM/F	100	0	0
25955 pJSP1A	LLMM/F	100	0.3	4
25955 pJSP1A	LLMM/F + I	100	0.15	2
25955 pJSP1A/pJSP2	LLMM/F	100	0.9	11
25955	LLMM/F + I	62	1.0	20

B. Conversion of various carbon substrates to 1,3-propanediol by *K. pneumoniae* (ATCC 25955 pJSP1A/pJSP2):

An aliquot (0.1 mL) of frozen stock cultures of *K. pneumoniae* (ATCC 25955 pJSP1A/pJSP2) was transferred to 50 mL Seed medium in a 250 mL baffled flask. The Seed medium contained per liter: 0.1 molar NaK/PO<sub>4</sub> buffer, pH 7.0; 3 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; 5 g glucose, 0.15 g MgSO<sub>4</sub>·7H<sub>2</sub>O, 10 mL 100X Trace Element solution, 25 mg chloramphenicol, 10 mg tetracycline, and 1 g yeast extract. The 100X Trace Element contained per liter: 10 g citric acid, 1.5 g CaCl<sub>2</sub>·2H<sub>2</sub>O, 2.8 g FeSO<sub>4</sub>·7H<sub>2</sub>O, 0.39 g ZnSO<sub>4</sub>·7H<sub>2</sub>O, 0.38 g CuSO<sub>4</sub>·5H<sub>2</sub>O, 0.2 g CoCl<sub>2</sub>·6H<sub>2</sub>O, and 0.3 g MnCl<sub>2</sub>·4H<sub>2</sub>O. The resulting solution was titrated to pH 7.0 with either KOH or H<sub>2</sub>SO<sub>4</sub>. The glucose, trace elements, antibiotics and yeast extracts were sterilized separately. The seed inoculum was grown overnight at 35 °C and 250 rpm.

The reaction design was semi-aerobic. The system consisted of 130 mL Reaction medium in 125 mL sealed flasks that were left partially open with aluminum foil strip. The Reaction Medium contained per liter: 3 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; 20 g carbon substrate; 0.15 molar NaK/PO<sub>4</sub> buffer, pH 7.5; 1 g yeast extract; 0.15 g MgSO<sub>4</sub>·7H<sub>2</sub>O; 0.5 mmoles IPTG; 10 mL 100X Trace Element solution; 25 mg chloramphenicol; and 10 mg tetracycline. The resulting solution was titrated to pH 7.5 with KOH or H<sub>2</sub>SO<sub>4</sub>. The carbon sources were: D-glucose (Glc); D-fructose (Frc); D-lactose (Lac); D-sucrose (Suc); D-maltose (Mal); and D-mannitol (Man). A few glass beads were included in the medium to improve mixing. The reactions were initiated by addition of seed inoculum so that the optical density of the cell suspension started at 0.1 AU as

-- 46 --

measured at 1600 nm. The flasks were incubated at 35 °C: 250 rpm. 3G production was measured by HPLC after 24 hr. Table 6 describes the yields of 1,3-propanediol produced from the various carbon substrates.

5

**Table 6**  
Production of 1,3-propanediol from various carbon substrates  
using recombinant *Klebsiella* 25955 pJSP1A/pJSP2

Carbon Substrate	1,3-Propanediol (g/L)		
	Expt. 1	Expt. 2	Expt 3
Glc	0.89	1	1.6
Frc	0.19	0.23	0.24
Lac	0.15	0.58	0.56
Suc	0.88	0.62	
Mal	0.05	0.03	0.02
Man	0.03	0.05	0.04

10

#### EXAMPLE 9

##### IMPROVEMENT OF 1,3-PROPANEDIOL PRODUCTION USING *dhaBX* GENE

Example 9 demonstrates the improved production of 1,3-propanediol in *E.coli* when a gene encoding a protein X is introduced.

##### Construction of expression vector pTacIQ

15

The *E. coli* expression vector, pTacIQ containing the lacIq gene (Farabaugh, P.J. 1978, Nature 274 (5673) 765-769) and tac promoter (Amann et al, 1983, Gene 25, 167-178) was inserted into the restriction endonuclease site EcoRI of pBR322 (Sutcliffe, 1979, Cold Spring Harb. Symp. Quant. Biol. 43, 77-90). A multiple cloning site and terminator sequence (SEQ ID NO:50) replaces the pBR322 sequence from EcoRI to SphI.

20

##### Subcloning the glycerol dehydratase genes (*dhaB1*, 2,3, X)

The region containing the entire coding region for *Klebsiella dhaB1*, *dhaB2*, *dhaB3* and *dhaBX* of the *dhaB* operon from pHK28-26 was cloning into pBluescriptIIKS+(Stratagene) using the restriction enzymes KpnI and EcoRI to create the plasmid pM7.

25

The open reading frame for *dhaB3* gene was amplified from pHK 28-26 by PCR using primers (SEQ ID NO:51 and SEQ ID NO:52) incorporating an EcoRI site at the 5' end and a XbaI site at the 3' end. The product was subcloned into pLitmus29(NEB) to generate the plasmid pDHAB3 containing *dhaB3*.

30

The *dhaBX* gene was removed by digesting plasmid pM7 with Apal and XbaI, purifying the 5.9 kb fragment and ligating it with the 325-bp Apal-XbaI fragment from plasmid pDHAB3 to create pM11 containing *dhaB1*, *dhaB2* and *dhaB3*.

-- 47 --

The open reading frame for the *dhaB1* gene was amplified from pHK28-26 by PCR using primers (SEQ ID NO:53 and SEQ ID NO:54) incorporating HindIII site and a consensus ribosome binding site at the 5' end and a XbaI site at the 3' end. The product was subcloned into pLitmus28(NEB) to generate the plasmids pDT1 containing *dhaB1*.

5 A NotI-XbaI fragment from pM11 containing part of the *dhaB1* gene, the *dhaB2* gene and the *dhaB3* gene with inserted into pDT1 to create the *dhaB* expression plasmid, pDT2. The HindIII-XbaI fragment containing the *dhaB(1,2,3)* genes from pDT2 was inserted into pTaciQ to create pDT3.

#### Subcloning the TMG dehydrogenase gene ( *dhaT* )

10 The KpnI-SacI fragment of pHK28-26, containing the TMG dehydrogenase (*dhaT*) gene, was subcloned into pBluescriptII KS+ creating plasmid pAH1. The *dhaT* gene was cloned by PCR from pAH1 as template DNA and synthetic primers (SEQ ID NO:55 with SEQ ID NO:56) incorporating an XbaI site at the 5' end and a BamHI site at the 3' end. The product was subcloned into pCR-Script(Stratagene) at the SrfI site to generate the plasmids pAH4 and pAH5  
15 containing *dhaT*. The pAH4 contains the *dhaT* gene in the right orientation for expression from the lac promoter in pCR-Script and pAH5 contains *dhaT* gene in the opposite orientation. The XbaI-BamHI fragment from pAH4 containing the *dhaT* gene was inserted into pTaciQ to generate plasmid, pAH8. The HindII-BamHI fragment from pAH8 containing the RBS and *dhaT* gene was inserted into pBluescriptIIKS+ to create pAH11.

#### Construction of an expression cassette for *dhaT* and *dhaB(1,2,3)*

20 An expression cassette for *dhaT* and *dhaB(1,2,3)* was assembled from the individual *dhaB(1,2,3)* and *dhaT* subclones described previously using standard molecular biology methods. A SpeI-SacI fragment containing the *dhaB(1,2,3)* genes from pDT3 was inserted into pAH11 at the SpeI-SacI sites to create pAH24. A Sall-XbaI linker (SEQ ID NO 57 and SEQ ID NO 58) was  
25 inserted into pAH5 which was digested with the restriction enzymes Sall-XbaI to create pDT16. The linker destroys the XbaI site. The 1 kb Sall-MluI fragment from pDT16 was then inserted into pAH24 replacing the existing Sall-MluI fragment to create pDT18.

#### Plasmid for the over-expression of *dhaT* and *dhaB(1, 2, 3, X)* in *E. coli*

30 The 4.4 kb NotI-XbaI fragment containing part of the *dhaB1* gene, *dhaB2*, *dhaB3* and *dhaBX* from plasmid pM7 was purified and ligated with the 4.1 Kb NotI-XbaI fragment from plasmid pDT18 (restoring *dhaB1*) to create pM33 containing the *dhaB1*, *dhaB2*, *dhaB3* and *dhaBX*.

#### *E. coli* strain

35 *E. coli* DH5a was obtained from BRL (Difco). This strain was transformed with the plasmids pM7, pM11, pM33 or pDT18 and selected on LA plates containing 100 ug/ml carbenicillin.

-- 48 --

Production of 1,3-propanediol

E. coli DH5a, containing plasmid pM7, pM11, pM33 or pDT18 was grown on LA plates plus 100 ug/ml carbenicillin overnight at 37°C. One colony from each was used to inoculate 25 ml of media (0.2 M KH<sub>2</sub>PO<sub>4</sub>, citric acid 2.0 g/L, MgSO<sub>4</sub>·7H<sub>2</sub>O 2.0 g/L, H<sub>2</sub>SO<sub>4</sub> (98%) 1.2 ml/L, Ferric ammonium citrate 0.3 g/L, CaCl<sub>2</sub>·2H<sub>2</sub>O 0.2 gram, yeast extract 5 g/L, glucose 10 g/L, glycerol 30 g/L,) plus Vitamine B12 0.005 g/L, 0.2 mM IPTG, 200 ug/ml carbenicillin and 5 ml modified Balch's trace-element solution (the composition of which can be found in Methods for General and Molecular Bacteriology (P. Gerhardt et al., eds, p 158, American Society for Microbiology, Washington, DC 1994), final pH 6.8 (NH<sub>4</sub>OH), then filter-sterilized in 250 ml erlenmeyers flasks. The shake flasks were incubated at 37°C with shaking (300 rpm) for several days, during which they were sampled for HPLC analysis by standard procedures. Final yields are shown in Table 4.

Overall, as shown in Table 7, the results indicate that the expression of dhaBX in plasmids expressing dhaB(1,2,3) or dhaT-dhaB(1,2,3) greatly enhances the production of 1,3-propanediol.

TABLE 7

Effect of *dhaBX* expression on the production of 1,3-propanediol by *E. coli*

Strain	Time (days)	1,3-propanediol (mg/L)*
DH5a/pM7 (dhaB1,2,3,X)	1	1500
	2	2700
DH5a/pM11 (dhaB1,2,3)	1	< 200 µg
	2	< 200 µg
DH5a/pM33 (dhaT-dhaB1,2,3,X)	2	1200
DH5a/pDT18 (dhaT-dhaB1,2,3)	2	88

\* Expressed as an average from several experiments.

Primers:

SEQ ID NO: 50- MCS-TERMINATOR:

5 AGCTTAGGAGTCTAGAATATTGAGCTCGAATTCCTGGGCATGCGGTACCGGATCCAGAAAA  
AAGCCCGCACCTGACAGTGCGGGCTTTTTTTTTT 3'

SEQ ID NO: 51 -dhaB3-5' end. EcoRI

GGAATTCAGATCTCAGCAATGAGCGAGAAAACCATGC

SEQ ID NO 52: dhaB3-3' end XbaI

-- 49 --

GCTCTAGATTAGCTTCCTTTACGCAGC

SEQ ID NO 53: *dhaB1* 5' end-HindIII-SD

5' GGCCAAGCTTAAGGAGGTTAATTAAATGAAAAG 3'

5

SEQ ID NO 54: *dhaB1* 3' end-XbaI

5' GCTCTAGATTATTCAATGGTGTCTGGG 3'

SEQ ID NO 55: *dhaT* 5' end-XbaI

10 5' GCGCCGTCTAGAATTATGAGCTATCGTATGTTTGATTATCTG 3'

SEQ ID NO 56: *dhaT* 3' end-BamHI

5' TCTGATACGGGATCCTCAGAATGCCTGGCGGAAAAT 3'

15 SEQ ID NO 57: pUSH Linker1:

5' TCGACGAATTCAGGAGGA 3'

SEQ ID NO 58: pUSH Linker2:

5' CTAGTCCTCCTGAATTCG 3'

20

#### EXAMPLE 10

##### Reactivation of the Glycerol Dehydratase Activity

Example 10 demonstrates the *in vivo* reactivation of the glycerol dehydratase activity in microorganisms containing at least one gene encoding protein X.

25 Plasmids pM7 and pM11 were constructed as described in Example 9 and transformed into *E. coli* DH5 $\alpha$  cells. The transformed cells were cultured and assayed for the production of 1,3-propanediol according to the method of Honda et al. (1980, In Situ Reactivation of Glycerol-Inactivated Coenzyme B<sub>12</sub>-Dependent Enzymes, Glycerol Dehydratase and Diol Dehydratase. Journal of Bacteriology 143:1458-1465).

##### 30 Materials and methods

###### Toluenization of Cells

The cells were grown to mid-log phase and were harvested by centrifugation at room temperature early in growth, i.e.  $0.2 > OD_{600} < 0.8$ . The harvested cells were washed 2x in 50mM KPO<sub>4</sub> pH8.0 at room temperature. The cells were resuspended to OD<sub>600</sub> 20-30 in 50mM KPO<sub>4</sub> pH8.0. The absolute OD is not critical. A lower cell mass is resuspended in less volume. If  
35 coenzyme B12 is added at this point, the remainder of the steps are performed in the dark.

-- 50 --

Toluene is added to 1% final volume of cell suspension and the suspension is shaken vigorously for 5 minutes at room temperature. The suspension is centrifuged to pellet the cells. The cells are washed 2x in 50mM KPO<sub>4</sub> pH8.0 at room temperature (25mls each). The cell pellet is resuspended in the same volume as was used prior to toluene addition and transfer to fresh  
 5 tubes. The OD<sub>600</sub> for the toluenized cells was measured and recorded and stored at 4 degrees C..

#### Whole Cell Glycerol Dehydratase Assay

The toluene treated cells were assayed at 37 degrees C for the presence of dehydratase activity. Three sets of reactions were carried out as shown below: no ATP, ATP added at 0 time, and ATP added at 10 minutes.

10	No ATP:	100ul	2M Glycerol
		100ul	150uM CoB <sub>12</sub>
		700ul	Buffer (0.03M KPO <sub>4</sub> / 0.5M KCl, pH8.0)
15	T=0 minute ATP	100ul	2M Glycerol
		100ul	150uM CoB <sub>12</sub>
		600ul	Buffer (0.03M KPO <sub>4</sub> / 0.5M KCl, pH8.0)
		100ul	30mM ATP/ 30mM MnCl <sub>2</sub>
20	T=10 minute ATP	100ul	2M Glycerol
		100ul	150uM CoB <sub>12</sub>
		700ul	Buffer (0.03M KPO <sub>4</sub> / 0.5M KCl, pH8.0)

Controls were prepared for each of the above conditions by adding 100uls buffer instead of CoB<sub>12</sub>. The tubes were mixed. 50uls MBTH (3-Methyl-2-Benzo-Thiazolinone Hydrazone) (6  
 25 mg/ml in 375mM Glycine / HCl pH2.7) was added to each of these tubes and continue incubation in ice water. The reaction tubes were placed in a 37 degree C water bath for a few minutes to equilibrate to 37 degree C. A tube containing enough toluenized cells for all assay tubes was placed into the 37 degree C water bath for a few minutes to equilibrate to 37 degree C. A tube containing 2.5 fold diluted (in assay buffer) 30mM ATP/ 30mM MnCl<sub>2</sub> (12mM each) was placed  
 30 into the 37 degree C water bath for a few minutes to equilibrate to 37 degree C. A 100ul cell suspension was added to all tubes and samples were taken at 0,1,2,3,4,5,10,15,20 and 30 minutes. At every timepoint, 100uls of reaction was withdrawn and immediately added to 50uls ice cold MBTH, vortexed, and placed in an ice water bath. At T=10 minutes, a sample was withdrawn and added to MBTH, then 100uls of the 2.5 fold diluted ATP/Mn was added as fast as  
 35 is possible. When all samples were collected, the sample tube rack was added to a boiling water bath and boiled for three minutes. The tubes were chilled in an ice water bath for 30 seconds.



-- 51 --

500ul of freshly prepared 3.3 mg/ml  $\text{FeCl}_3 \cdot 6\text{H}_2\text{O}$ , was added to the tubes and the tubes vortexed. The tubes were incubated at room temperature for 30 minutes, diluted 10x in  $\text{H}_2\text{O}$ , and then centrifuged to collect the cells and particulates. The absorbance was measured at 670nm and the cells were diluted to keep OD under 1.0.

#### 5      Example of Calculation of Activity

The observed OD670 was multiplied by the dilution factor to determine absorbance. The blank absorbance was subtracted for that reaction series and the T0 A670nm was subtracted. The absolute A670nm was divided by 53.4 (mM extinction coefficient for 3OH-propionaldehyde) and the mM concentration was multiplied by any dilution of reaction during timecourse. Because 1 ml  
10 reaction was used, the concentration (umoles/ml) of 3OH-propionaldehyde was divided by the mgs dry weight used in the assay (calculated via OD600 and  $1\text{OD } 600 = 0.436 \text{ mgs dry weight}$ ) to get umoles aldehyde per mg dry weight cells.

#### Results

15      As shown in Figure 6, whole E.coli cells were assayed for reactivation of glycerol dehydratase in the absence and presence of added ATP and  $\text{Mn}^{++}$ . The results indicate that cells containing a plasmid carrying dhaB 1, 2 and 3 as well as protein X have the ability to reactivate catalytically inactivated glycerol dehydrogenase. Cells containing protein 1, protein 2 and protein 3 have increased ability to reactivate the catalytically inactivated glycerol dehydratase.

20      As shown in Figure 7, whole E.coli cells were assayed for reactivation of glycerol-inactivated glycerol dehydratase in the absence and in the presence of added ATP and  $\text{Mn}^{++}$ . The results show that cells containing dhaB subunits 1, 2 and 3 and X have the ability to reactivate catalytically inactivated glycerol dehydratase. Cell lacking the protein X gene do not have the ability to reactivate the catalytically inactivated glycerol dehydratase.

25      Figures 9 and 10 illustrate that host cells containing plasmid pHK 28-26 (Figure 1), when cultured under conditions suitable for the production of 1,3-propanediol, produced more 1,3-propanediol than host cells transformed with pDT24 and cultured under conditions suitable for the production of 1,3-propanediol. Plasmid pDT24 is a derivative of pDT18 (described in Example 9) and contains dhaT, dhaB 1, 2, 3 and protein X, but lacks proteins 1, 2 and 3.

-- 52 --

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- (i) APPLICANT: MARIA DIAZ-TORRES  
NIGEL DUNN-COLEMAN  
MATTHEW CHASE
- (ii) TITLE OF INVENTION: METHOD FOR THE  
RECOMBINANT PRODUCTION OF 1,3 PROPANEDIOL
- (iii) NUMBER OF SEQUENCES: 49
- (iv) CORRESPONDENCE ADDRESS:
  - (A) ADDRESSEE: GENENCOR INTERNATIONAL, INC.
  - (B) STREET: 4 CAMBRIDGE PLACE  
1870 SOUTH WINTON ROAD
  - (C) CITY: ROCHESTER
  - (D) STATE: NEW YORK
  - (E) COUNTRY: U.S.A.
  - (F) POSTAL CODE (ZIP): 14618
- (v) COMPUTER READABLE FORM:
  - (A) MEDIUM TYPE: 3.50 INCH DISKETTE
  - (B) COMPUTER: IBM PC COMPATIBLE
  - (C) OPERATING SYSTEM: MICROSOFT WINDOWS 3.1
  - (D) SOFTWARE: MICROSOFT WORD 2.0C
- (vi) CURRENT APPLICATION DATA:
  - (A) APPLICATION NUMBER:
  - (B) FILING DATE: 11/13/97
  - (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
  - (A) APPLICATION NUMBER: 60/030,601
  - (B) FILING DATE: 11/13/96
  - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
  - (A) NAME: GLAISTER, DEBRA
  - (B) REGISTRATION NO.: 33,888
  - (C) REFERENCE/DOCKET NUMBER: GC 369-2
  - (ix) TELECOMMUNICATION INFORMATION:
    - (A) TELEPHONE: 650-864-7620
    - (B) TELEFAX: 650-845-6504

## (2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1668 base pairs

-- 53 --

(B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: DHAB1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATGAAAAGAT CAAAACGATT TGCAGTACTG GCCCAGCGCC CCGTCAATCA GGACGGGCTG	60
ATTGGCGAGT GGCCTGAAGA GGGGCTGATC GCCATGGACA GCCCCTTTGA CCCGGTCTCT	120
TCAGTAAAAG TGGACAACGG TCTGATCGTC GAACTGGACG GCAAACGCCG GGACCAGTTT	180
GACATGATCG ACCGATTAT CGCCGATTAC GCGATCAACG TTGAGCGCAC AGAGCAGGCA	240
ATGCGCCTGG AGGCGGTGGA AATAGCCCGT ATGCTGGTGG ATATTCACGT CAGCCGGGAG	300
GAGATCATTG CCATCACTAC CGCCATCACG CCGGCCAAAG CGGTCGAGGT GATGGCGCAG	360
ATGAACGTGG TGGAGATGAT GATGGCGCTG CAGAAGATGC GTGCCCCCGG GACCCCTCC	420
AACCAGTGCC ACGTCACCAA TCTCAAAGAT AATCCGGTGC AGATTGCCGC TGACGCCGCC	480
GAGGCCGGGA TCCGCGGCTT CTCAGAACAG GAGACCACGG TCGGTATCGC GCGCTACGCG	540
CCGTTTAACG CCCTGGCGCT GTTGGTCGGT TCGCAGTGCG GCCGCCCCGG CGTGTGACG	600
CAGTGCTCGG TGGAAGAGGC CACCGAGCTG GAGCTGGGCA TCGGTGGCTT AACCAGCTAC	660
GCCGAGACGG TGTGGTCTA CGGCACCGAA GCGGTATTTA CCGACGGCGA TGATACGCCG	720
TGGTCAAAGG CGTTCCTCGC CTCGGCCTAC GCCTCCCGCG GGTGAAAAT GCGCTACACC	780
TCCGGCACCG GATCCGAAGC GCTGATGGGC TATTCGGAGA GCAAGTCGAT GCTCTACCTC	840
GAATCGCGCT GCATCTTCAT TACTAAAGGC GCCGGGGTTC AGGGACTGCA AAACGGCGCG	900
GTGAGCTGTA TCGGCATGAC CGGCGCTGTG CCGTCGGGCA TTCGGGCGGT GCTGGCGGAA	960
AACCTGATCG CCTCTATGCT CGACCTCGAA GTGGCGTCCG CCAACGACCA GACTTTCTCC	1020
CACTCGGATA TTCGCCGCAC CGCGCGCACC CTGATGCAGA TGCTGCCGGG CACCGACTTT	1080
ATTTTCTCCG GCTACAGCGC GGTGCCGAAC TACGACAACA TGTTCGCCGG CTCGAACTTC	1140
GATGCGGAAG ATTTTGATGA TTACAACATC CTGCAGCGTG ACCTGATGGT TGACGGCGGC	1200
CTGCGTCCGG TGACCGAGGC GGAAACCATT GCCATTGCC AGAAAGCGGC GCGGGCGATC	1260

-- 54 --

CAGGCGGTTT TCCGCGAGCT GGGGCTGCCG CCAATCGCCG ACGAGGAGGT GGAGGCCGCC 1320  
 ACCTACGCGC ACGGCAGCAA CGAGATGCCG CCGCGTAACG TGGTGGAGGA TCTGAGTGCG 1380  
 GTGGAAGAGA TGATGAAGCG CAACATCACC GGCCTCGATA TTGTCGGCGC GCTGAGCCGC 1440  
 AGCGGCTTTG AGGATATCGC CAGCAATATT CTCAATATGC TGCGCCAGCG GGTACCCGGC 1500  
 GATTACCTGC AGACCTCGGC CATTCTCGAT CGGCAGTTCG AGGTGGTGAG TCGGCTCAAC 1560  
 GACATCAATG ACTATCAGGG GCCGGGCACC GGCTATCGCA TCTCTGCCGA ACGCTGGGCG 1620  
 GAGATCAAAA ATATTCCGGG CQTGGTTCAG CCCGACACCA TTGAATAA 1668

## (2) INFORMATION FOR SEQ ID NO:2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 585 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: DHAB2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GTGCAACAGA CAACCCAAAT TCAGCCCTCT TTTACCCTGA AAACCCGCGA GGGCGGGGTA 60  
 GCTTCTGCCG ATGAACGCGC CGATGAAGTG GTGATCGGCG TCGGCCCTGC CTTTCGATAAA 120  
 CACCAGCATC AACTCTGAT CGATATGCCC CATGGCGCGA TCCTCAAAGA GCTGATTGCC 180  
 GGGGTGGAAG AAGAGGGGCT TCACGCCCCG GTGGTGCACA TTCTGCGCAC GTCCGACGTC 240  
 TCCTTTATGG CCTGGGATGC GGCACACCTG AGCGGCTCGG GGATCGGCAT CGGTATCCAG 300  
 TCGAAGGGGA CCACGGTCAT CCATCAGCGC GATCTGCTGC CGCTCAGCAA CCTGGAGCTG 360  
 TTCTCCCAGG CGCCGCTGCT GACGCTGGAG ACCTACCGGC AGATTGGCAA AACGCTGCG 420  
 CGCTATGCGC GCAAAGAGTC ACCTTCGCCG GTGCCGGTGG TGAACGATCA GATGGTGCGG 480  
 CCGAAATTTA TGGCCAAAGC CGCGCTATTT CATATCAAAG AGACCAAACA TGTGGTGCAG 540  
 GACGCCGAGC CCGTCACCCT GCACATCGAC TTAGTAAGGG AGTGA 585

## (2) INFORMATION FOR SEQ ID NO:3:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 426 base pairs
- (B) TYPE: nucleic acid

-- 55 --

- (C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: DHAB3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

ATGAGCGAGA AAACCATGCG CGTGCAGGAT TATCCGTTAG CCACCCGCTG CCCGGAGCAT    60
ATCCTGACGC CTACCGGCAA ACCATTGACC GATATTACCC TCGAGAAGGT GCTCTCTGGC    120
GAGGTGGGCC CGCAGGATGT GCGGATCTCC CGCCAGACCC TTGAGTACCA GGCGCAGATT    180
GCCGAGCAGA TGCAGCGCCA TGCGGTGGCG CGCAATTTCC GCCGCGCGGC GGAGCTTATC    240
GCCATTCCTG ACGAGCGCAT TCTGGCTATC TATAACGCGC TGCGCCCGTT CCGCTCCTCG    300
CAGGCGGAGC TGCTGGCGAT CGCCGACGAG CTGGAGCACA CCTGGCATGC GACAGTGAAT    360
GCCGCCTTTG TCCGGGAGTC GGC GGAAGTG TATCAGCAGC GGCATAAGCT GCGTAAAGGA    420
AGCTAA                                           426

```

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1164 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: DHAT

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```

ATGAGCTATC GTATGTTTGA TTATCTGGTG CCAAACGTTA ACTTTTTTTGG CCCCAACGCC    60
ATTTCCGTAG TCGGCGAACG CTGCCAGCTG CTGGGGGGGA AAAAAGCCCT GCTGGTCACC    120
GACAAAGGCC TGCGGGCAAT TAAAGATGGC GCGGTGGACA AAACCCTGCA TTATCTGCGG    180
GAGGCCGGGA TCGAGGTGGC GATCTTTGAC GCGGTCGAGC CGAACCCGAA AGACACCAAC    240
GTGCGCGACG GCCTCGCCGT GTTTCGCCGC GAACAGTGCG ACATCATCGT CACCGTGGGC    300
GGCGGCAGCC CGCACGATTG CGGCAAAGGC ATCGGCATCG CCGCCACCCA TGAGGGCGAT    360
CTGTACCACT ATGCCGGAAT CGAGACCCCTG ACCAACCCGC TGCCGCCTAT CGTCGCGGTC    420
AATACCACCG CCGGCACCGC CAGCGAGGTC ACCCGCCACT GCGTCCTGAC CAACACCGAA    480

```

-- 56 --

ACCAAAGTGA AGTTTGTGAT CGTCAGCTGG CGCAAACCTGC CGTCGGTCTC TATCAACGAT 540  
 CCACTGCTGA TGATCGGTAA ACCGGCCGCC CTGACCGCGG CGACCGGGAT GGATGCCCTG 600  
 ACCCAGCCCG TAGAGGCCTA TATCTCCAAA GACGCTAACC CGGTGACGGA CGCCGCCGCC 660  
 ATGCAGGCGA TCCGCCTCAT CGCCCACAAC CTGCGCCAGG CCGTGGCCCT CGGCAGCAAT 720  
 CTGCAGGCGC GGGAAAACAT GGCCTATGCT TCTCTGCTGG CCGGGATGGC TTTCAATAAC 780  
 GCCAACCTCG GCTACGTGCA CGCCATGGCG CACCAGCTGG GCGGCCTGTA CGACATGCCG 840  
 CACGGCGTGG CCAACGCTGT CCTGCTGCCG CATGTGGCGC GCTACAACT GATCGCCAAC 900  
 CCGGAGAAAT TCGCCGATAT CGCTGAACTG ATGGGCGAAA ATATCACCGG ACTGTCCACT 960  
 CTCGACGCGG CGGAAAAAGC CATCGCCGCT ATCACGCGTC TGTCGATGGA TATCGGTATT 1020  
 CCGCAGCATC TGCGCGATCT GGGGGTAAAA GAGGCCGACT TCCCCTACAT GCGGAGATG 1080  
 GCTCTAAAAG ACGGCAATGC GTTCTCGAAC CCGCGTAAAG GCAACGAGCA GGAGATTGCC 1140  
 GCGATTTTCC GCCAGGCATT CTGA 1164

## (2) INFORMATION FOR SEQ ID NO:5:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1380 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPD1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

CTTTAATTTT CTTTATCTT ACTCTCTAC ATAAGACATC AAGAAACAAT TGTATATTGT 60  
 ACACCCCCCT CCTCCACAAA CACAAATATT GATAATATAA AGATGTCTGC TGCTGCTGAT 120  
 AGATTAAACT TAACTTCCGG CCACTTGAAT GCTGGTAGAA AGAGAAGTTC CTCTTCTGTT 180  
 TCTTTGAAGG CTGCCGAAAA GCCTTTCAAG GTTACTGTGA TTGGATCTGG TAACTGGGGT 240  
 ACTACTATTG CCAAGGTGGT TGCCGAAAAT TGTAAGGGAT ACCCAGAAGT TTTGCTCCA 300  
 ATAGTACAAA TGTGGGTGTT CGAAGAAGAG ATCAATGGTG AAAAAATTGAC TGAAATCATA 360  
 AATACTAGAC ATCAAACGT GAAATACTTG CCTGGCATCA CTCTACCCGA CAATTTGGTT 420  
 GCTAATCCAG ACTTGATTGA TTCAGTCAAG GATGTCGACA TCATCGTTTT CAACATTCCA 480

-- 57 --

CATCAATTTT TGCCCCGTAT CTGTAGCCAA TTGAAAGGTC ATGTTGATTC ACACGTCAGA 540  
 GCTATCTCCT GTCTAAAGGG TTTTGAAGTT GGTGCTAAAG GTGTCCAATT GCTATCCTCT 600  
 TACATCACTG AGGAACTAGG TATTCAATGT GGTGCTCTAT CTGGTGCTAA CATTGCCACC 660  
 GAAGTCGCTC AAGAACACTG GTCTGAAACA ACAGTTGCTT ACCACATTCC AAAGGATTTC 720  
 AGAGGCGAGG GCAAGGACGT CGACCATAAG GTTCTAAAGG CCTTGTTCCTA CAGACCTTAC 780  
 TTCCACGTTA GTGTCATCGA AGATGTTGCT GGTATCTCCA TCTGTGGTGC TTTGAAGAAC 840  
 GTTGTTCGCT TAGGTTGTGG TTTCGTCGAA GGTCTAGGCT GGGGTAACAA CGCTTCGCT 900  
 GCCATCCAAA GAGTCGGTTT GGGTGAGATC ATCAGATTCTG GTCAAATGTT TTTCCCAGAA 960  
 TCTAGAGAAG AAACATACTA CCAAGAGTCT GCTGGTGTG CTGATTGAT CACCACCTGC 1020  
 GCTGGTGGTA GAAACGTCAA GGTGCTAGG CTAATGGCTA CTTCTGGTAA GGACGCCTGG 1080  
 GAATGTGAAA AGGAGTTGTT GAATGGCCAA TCCGCTCAAG GTTTAATTAC CTGCAAAGAA 1140  
 GTTCACGAAT GGTGGAAC ATGTGGCTCT GTCGAAGACT TCCCATATT TGAAGCCGTA 1200  
 TACCAAATCG TTTACAACAA CTACCCAATG AGAACCTGC CGGACATGAT TGAAGAATTA 1260  
 GATCTACATG AAGATTAGAT TTATTGGAGA AAGATAACAT ATCATACTTC CCCCCTTTT 1320  
 TTCGAGGCTC TTCTATATCA TATTCATAAA TTAGCATTAT GTCATTCTC ATAACTACTT 1380

## (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2946 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: GPD2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GAATTCGAGC CTGAAGTGCT GATTACCTTC AGGTAGACTT CATCTTGACC CATCAACCCC 60  
 AGCGTCAATC CTGCAAATAC ACCACCCAGC AGCACTAGGA TGATAGAGAT AATATAGTAC 120  
 GTGGTAACGC TTGCCTCATC ACCTACGCTA TGGCCGGAAT CGGCAACATC CCTAGAATTG 180  
 AGTACGTGTG ATCCGGATAA CAACGGCAGT GAATATATCT TCGGTATCGT AAAGATGTGA 240  
 TATAAGATGA TGTATACCCA ATGAGGAGCG CCTGATCGTG ACCTAGACCT TAGTGGCAAA 300  
 AACGACATAT CTATTATAGT GGGGAGAGTT TCGTGCAAAT AACAGACGCA GCAGCAAGTA 360

-- 58 --

ACTGTGACGA TATCAACTCT TTTTTTATTA TGTAATAAGC AAACAAGCAC GAATGGGGAA 420  
AGCCTATGTG CAATCACCAA GGTCGTCCCT TTTTTCCTAT TTGCTAATTT AGAATTTAAA 480  
GAAACCAAAA GAATGAAGAA AGAAAACAAA TACTAGCCCT AACCTGACT TCGTTTCTAT 540  
GATAATACCC TGCTTTAATG AACGGTATGC CCTAGGGTAT ATCTCACTCT GTACGTTACA 600  
AACTCCGGTT ATTTTATCGG AACATCCGAG CACCCGCGCC TTCCTCAACC CAGGCACCGC 660  
CCCAGGTAAC CGTGCGCGAT GAGCTAATCC TGAGCCATCA CCCACCCAC CCGTTGATGA 720  
CAGCAATTCTG GGAGGGCGAA AATAAACTG GAGCAAGGAA TTACCATCAC CGTCACCATC 780  
ACCATCATAT CGCCTTAGCC TCTAGCCATA GCCATCATGC AAGCGTGTAT CTTCTAAGAT 840  
TCAGTCATCA TCATTACCGA GTTGTTTTTC CTTACATGA TGAAGAAGGT TTGAGTATGC 900  
TCGAAACAAT AAGACGACGA TGGCTCTGCC ATTGGTTATA TTACGCTTTT GCGGCGAGGT 960  
GCCGATGGGT TGCTGAGGGG AAGAGTGTTT AGCTTACGGA CCTATTGCCA TTGTTATTCC 1020  
GATTAATCTA TTGTTTACGA GCTCTTCTCT ACCCTGTCAT TCTAGTATTT TTTTTTTTTT 1080  
TTTTTGTTTT TACTTTTTTT TCTTCTTGCC TTTTTTCTT GTTACTTTTT TTCTAGTTTT 1140  
TTTTCTTCC ACTAAGCTTT TTCCTTGATT TATCCTGGG TTCTTCTTC TACTCCTTTA 1200  
GATTTTTTTT TTATATATTA ATTTTAAAGT TTATGTATTT TGGTAGATTC AATTCTCTTT 1260  
CCCTTTCCTT TTCCTTCGCT CCCCTTCCTT ATCAATGCTT GCTGTCAGAA GATTAACAAG 1320  
ATACACATTC CTTAAGCGAA CGCATCCGGT GTTATATACT CGTCGTGCAT ATAAAATTTT 1380  
GCCTTCAAGA TCTACTTTCC TAAGAAGATC ATTATTACAA ACACAACTGC ACTCAAAGAT 1440  
GACTGCTCAT ACTAATATCA AACAGCACAA AACTGTGCAT GAGGACCATC CTATCAGAAG 1500  
ATCGGACTCT GCCGTGTCAA TTGTACATTT GAAACGTGCG CCCTTCAAGG TTACAGTGAT 1560  
TGGTCTGGT AACTGGGGGA CCACCATCGC CAAAGTCATT GCGGAAAACA CAGAATTGCA 1620  
TTCCCATATC TTCGAGCCAG AGGTGAGAAT GTGGGTTTTT GATGAAAAGA TCGGCGACGA 1680  
AAATCTGACG GATATCATAA ATACAAGACA CCAGAACGTT AAATATCTAC CCAATATTGA 1740  
CCTGCCCCAT AATCTAGTGG CCGATCCTGA TCTTTTACAC TCCATCAAGG GTGCTGACAT 1800  
CCTTGTTTTT AACATCCCTC ATCAATTTTT ACCAAACATA GTCAAACAAT TGCAAGGCCA 1860  
CGTGGCCCTT CATGTAAGGG CCATCTCGTG TCTAAAAGGG TTCGAGTTGG GCTCCAAGGG 1920  
TGTGCAATTG CTATCCTCCT ATGTTACTGA TGAGTTAGGA ATCCAATGTG GCGCACTATC 1980  
TGGTGAAAC TTGGCACCGG AAGTGGCCAA GGAGCATTTG TCCGAAACCA CCGTGGCTTA 2040



-- 59 --

CCAACTACCA AAGGATTATC AAGGTGATGG CAAGGATGTA GATCATAAGA TTTTGAAATT 2100  
 GCTGTTCCAC AGACCTTACT TCCACGTCAA TGTCATCGAT GATGTTGCTG GTATATCCAT 2160  
 TGCCGGTGCC TTGAAGAACG TCGTGGCACT TGCATGTGGT TTCGTAGAAG GTATGGGATG 2220  
 GGGTAACAAT GCCTCCGCAG CCATTCAAAG GCTGGGTTTA GGTGAAATTA TCAAGTTCGG 2280  
 TAGAATGTTT TTCCCAGAAT CCAAAGTCGA GACCTACTAT CAAGAATCCG CTGGTGTTGC 2340  
 AGATCTGATC ACCACCTGCT CAGGCGGTAG AAACGTCAAG GTTGCCACAT ACATGGCCAA 2400  
 GACCGGTAAG TCAGCCTTGG AAGCAGAAAA GGAATTGCTT AACGGTCAAT CCGCCCAAGG 2460  
 GATAATCACA TGCAGAGAAG TTCACGAGTG GCTACAAACA TGTGAGTTGA CCCAAGAATT 2520  
 CCCAATTATT CGAGGCAGTC TACCAGATAG TCTACAACAA CGTCCGCATG GAAGACCTAC 2580  
 CGGAGATGAT TGAAGAGCTA GACATCGATG ACGAATAGAC ACTCTCCCCC CCCCTCCCCC 2640  
 TCTGATCTTT CCTGTTGCCT CTTTTTCCCC CAACCAATTT ATCATTATAC ACAAGTTCTA 2700  
 CAACTACTAC TAGTAACATT ACTACAGTTA TTATAATTTT CTATTCTCTT TTTCTTTAAG 2760  
 AATCTATCAT TAACGTTAAT TTCTATATAT ACATAACTAC CATTATACAC GCTATTATCG 2820  
 TTTACATATC ACATCACCGT TAATGAAAGA TACGACACCC TGTACACTAA CACAATTTAA 2880  
 TAATCGCCAT AACCTTTTCT GTTATCTATA GCCCTTAAAG CTGTTTCTTC GAGCTTTTCA 2940  
 CTGCAG 2946

## (2) INFORMATION FOR SEQ ID NO:7:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3178 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: GUT2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CTGCAGAACT TCGTCTGCTC TGTGCCCATC CTCGCGGTTA GAAAGAAGCT GAATTGTTTC 60  
 ATGCGCAAGG GCATCAGCGA GTGACCAATA ATCACTGCAC TAATTCCTTT TTAGCAACAC 120  
 ATACTTATAT ACAGCACCAG ACCTTATGTC TTTTCTCTGC TCCGATACGT TATCCCACCC 180  
 AACTTTTATT TCAGTTTGG CAGGGGAAAT TTCACAACCC CGCACGCTAA AAATCGTATT 240

-- 60 --

TAAACTTAAA	AGAGAACAGC	CACAAATAGG	GAAC TTTGGT	CTAAACGAAG	GACTCTCCCT	300
CCCTTATCTT	GACCGTGCTA	TTGCCATCAC	TGCTACAAGA	CTAAATACGT	ACTAATATAT	360
GTTTTCGGTA	ACGAGAAGAA	GAGCTGCCGG	TGCAGCTGCT	GCCATGGCCA	CAGCCACGGG	420
GACGCTGTAC	TGGATGACTA	GCCAAGGTGA	TAGGCCGTTA	GTGCACAATG	ACCCGAGCTA	480
CATGGTGCAA	TTCCCCACCG	CCGCTCCACC	GGCAGGTCTC	TAGACGAGAC	CTGCTGGACC	540
GTCTGGACAA	GACGCATCAA	TTCGACGTGT	TGATCATCGG	TGGCGGGGCC	ACGGGGACAG	600
GATGTGCCCT	AGATGCTGCG	ACCAGGGGAC	TCAATGTGGC	CCTTGTTGAA	AAGGGGGATT	660
TTGCCTCGGG	AACGTCGTCC	AAATCTACCA	AGATGATTCA	CGGTGGGGTG	CGGTACTTAG	720
AGAAGGCCTT	CTGGGAGTTC	TCCAAGGCAC	AACTGGATCT	GGTCATCGAG	GCACTCAACG	780
AGCGTAAACA	TCTTATCAAC	ACTGCCCTC	ACCTGTGCAC	GGTGCTACCA	ATTCTGATCC	840
CCATCTACAG	CACCTGGCAG	GTCCCGTACA	TCTATATGGG	CTGTAAATTC	TACGATTTCT	900
TTGGCGGTTC	CCAAACTTG	AAAAATCAT	ACCTACTGTC	CAAATCCGCC	ACCGTGGAGA	960
AGGCTCCCAT	GCTTACCACA	GACAATTTAA	AGGCCTCGCT	TGTGTACCAT	GATGGGTCCT	1020
TTAACGACTC	GCGTTTGAAC	GCCACTTTAG	CCATCACGGG	TGTGGAGAAC	GGCGCTACCG	1080
TCTTGATCTA	TGTCGAGGTA	CAAAAATTGA	TCAAAGACCC	AACTTCTGGT	AAGGTTATCG	1140
GTGCCGAGGC	CCGGGACGTT	GAGACTAATG	AGCTTGTGAG	AATCAACGCT	AAATGTGTGG	1200
TCAATGCCAC	GGGCCCATAC	AGTGACGCCA	TTTGTCAAAT	GGACCGCAAC	CCATCCGGTC	1260
TGCCGGACTC	CCCCTAAAC	GACAACTCCA	AGATCAAGTC	GACTTTCAAT	CAAATCTCCG	1320
TCATGGACCC	GAAAATGGTC	ATCCCATCTA	TTGGCGTTCA	CATCGTATTG	CCCTCTTTTT	1380
ACTCCCCGAA	GGATATGGGT	TTGTTGGACG	TCAGAACCTC	TGATGGCAGA	GTGATGTTCT	1440
TTTTACCTTG	GCAGGGCAAA	GTCCTTGCCG	GCACCACAGA	CATCCCACTA	AAGCAAGTCC	1500
CAGAAAACCC	TATGCCTACA	GAGGCTGATA	TTCAAGATAT	CTTGAAAGAA	CTACAGCACT	1560
ATATCGAATT	CCCCGTGAAA	AGAGAAGACG	TGCTAAGTGC	ATGGGCTGGT	GTCAGACCTT	1620
TGGTCAGAGA	TCCACGTACA	ATCCCCGCAG	ACGGGAAGAA	GGGCTCTGCC	ACTCAGGGCG	1680
TGGTAAGATC	CCACTTCTTG	TTCACTTCGG	ATAATGGCCT	AATTACTATT	GCAGGTGGTA	1740
AATGGACTAC	TTACAGACAA	ATGGCTGAGG	AAACAGTCGA	CAAAGTTGTC	GAAGTTGGCG	1800
GATTCCACAA	CCTGAAACCT	TGTCACACAA	GAGATATTAA	GCTTGCTGGT	GCAGAAGAAT	1860
GGACGCAAAA	CTATGTGGCT	TTATTGGCTC	AAAAC TACCA	TTTATCATCA	AAAATGTCCA	1920

-- 61 --

ACTACTTGGT TCAAACTAC GGAACCCGTT CCTCTATCAT TTGCGAATTT TTCAAAGAAT 1980  
 CCATGGAAAA TAAACTGCCT TTGTCCTTAG CCGACAAGGA AAATAACGTA ATCTACTCTA 2040  
 GCGAGGAGAA CAACTTGGTC AATTTTGATA CTTTCAGATA TCCATTCCACA ATCGGTGAGT 2100  
 TAAAGTATTC CATGCAGTAC GAATATTGTA GAACTCCCTT GGACTTCCTT TTAAGAAGAA 2160  
 CAAGATTCGC CTTCTTGGAC GCCAAGGAAG CTTTGAATGC CGTGCATGCC ACCGTCAAAG 2220  
 TTATGGGTGA TGAGTTCAAT TGGTCGGAGA AAAAGAGGCA GTGGGAACTT GAAAAAAGT 2280  
 TGAAC TTCAT CCAAGGACGT TTCGGTGTCT AAATCGATCA TGATAGTTAA GGGTGACAAA 2340  
 GATAACATTC ACAAGAGTAA TAATAATGGT AATGATGATA ATAATAATAA TGATAGTAAT 2400  
 AACAATAATA ATAATGGTGG TAATGGCAAT GAAATCGCTA TTATTACCTA TTTTCCTTAA 2460  
 TGGAAGAGTT AAAGTAACT AAAAAAAGT CAAAAATATA TGAAGAAAAA AAAAAAAGA 2520  
 GGTAATAGAC TCTACTACTA CAATTGATCT TCAAATTATG ACCTTCCTAG TGTTTATATT 2580  
 CTATTTCCAA TACATAATAT AATCTATATA ATCATTGCTG GTAGACTTCC GTTTTAATAT 2640  
 CGTTTTAATT ATCCCCTTTA TCTCTAGTCT AGTTTTATCA TAAAATATAG AAACACTAAA 2700  
 TAATATTCTT CAAACGGTCC TGGTGCATAC GCAATACATA TTTATGGTGC AAAAAAAAAA 2760  
 ATGGAAAATT TTGCTAGTCA TAAACCCCTT CATAAAACAA TACGTAGACA TCGCTACTTG 2820  
 AAATTTTCAA GTTTTTATCA GATCCATGTT TCCTATCTGC CTTGACAACC TCATCGTCGA 2880  
 AATAGTACCA TTTAGAACGC CCAATATTCA CATTGTGTTC AAGGTCTTTA TTCACCAGTG 2940  
 ACGTGTAATG GCCATGATTA ATGTGCCTGT ATGGTTAACC ACTCCAAATA GCTTATATTT 3000  
 CATAGTGTC TGTGTTTTCA ATATAATGTT TAGTATCAAT GGATATGTTA CGACGGTGTT 3060  
 ATTTTCTTG GTCAAATCGT AATAAAATCT CGATAAATGG ATGACTAAGA TTTTGGTAA 3120  
 AGTTACAAAA TTTATCGTTT TCACTGTTGT CAATTTTTTG TTCTTGTAAT CACTCGAG 3178

## (2) INFORMATION FOR SEQ ID NO:8:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 816 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPPI

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

-- 62 --

```

ATGAAACGTT TCAATGTTTT AAAATATATC AGAACAAACA AAGCAAATAT ACAAACCATC    60
GCAATGCCTT TGACCACAAA ACCTTTATCT TTGAAAATCA ACGCCGCTCT ATTCGATGTT   120
GACGGTACCA TCATCATCTC TCAACCAGCC ATTGCTGCTT TCTGGAGAGA TTTCCGTAAA   180
GACAAGCCTT ACTTCGATGC CGAACACGTT ATTCACATCT CTCACGGTTG GAGAACTTAC   240
GATGCCATTG CCAAGTTCGC TCCAGACTTT GCTGATGAAG AATACGTTAA CAAGCTAGAA   300
GGTGAAATCC CAGAAAAGTA CGGTGAACAC TCCATCGAAG TTCCAGGTGC TGTCAAGTTG   360
TGTAATGCTT TGAACGCCTT GCCAAAGGAA AAATGGGCTG TCGCCACCTC TGGTACCCGT   420
GACATGGCCA AGAAATGGTT CGACATTTTG AAGATCAAGA GACCAGAATA CTTTCATCACC   480
GCCAATGATG TCAAGCAAGG TAAGCCTCAC CCAGAACCAT ACTTAAAGGG TAGAAACGGT   540
TTGGGTTTCC CAATTAATGA ACAAGACCCA TCCAAATCTA AGGTTGTTGT CTTTGAAGAC   600
GCACCAGCTG GTATTGCTGC TGGTAAGGCT GCTGGCTGTA AAATCGTTGG TATTGCTACC   660
ACTTTCGATT TGGACTTCTT GAAGGAAAAG GGTGTGACA TCATTGTCAA GAACCACGAA   720
TCTATCAGAG TCGGTGAATA CAACGCTGAA ACCGATGAAG TCGAATTGAT CTTTGATGAC   780
TACTTATACG CTAAGGATGA CTGTTGAAA TGGTAA                                816

```

## (2) INFORMATION FOR SEQ ID NO:9:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 753 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPP2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```

ATGGGATTGA CTAATAAACC TCTATCTTTG AAAGTTAACG CCGCTTTGTT CGACGTCGAC    60
GGTACCATTA TCATCTCTCA ACCAGCCATT GCTGCATTCT GGAGGGATT TCGGTAAGGAC   120
AAACCTTATT TCGATGCTGA ACACGTTATC CAAGTCTCGC ATGGTTGGAG AACGTTTGAT   180
GCCATTGCTA AGTTCGCTCC AGACTTTGCC AATGAAGAGT ATGTTAACAA ATTAGAAGCT   240
GAAATTCCGG TCAAGTACGG TGAATAATCC ATTGAAGTCC CAGGTGCAGT TAAGCTGTGC   300
AACGCTTTGA ACGCTCTACC AAAAGAGAAA TGGGCTGTGG CAACTCCGG TACCCGTGAT   360

```

-- 63 --

ATGGCACAAA AATGGTTCGA GCATCTGGGA ATCAGGAGAC CAAAGTACTT CATTACCGCT 420  
 AATGATGTCA AACAGGGTAA GCCTCATCCA GAACCATATC TGAAGGGCAG GAATGGCTTA 480  
 GGATATCCGA TCAATGAGCA AGACCCTTCC AAATCTAAGG TAGTAGTATT TGAAGACGCT 540  
 CCAGCAGGTA TTGCCGCCGG AAAAGCCGCC GGTGTGAAGA TCATTGGTAT TGCCACTACT 600  
 TTCGACTTGG ACTTCCTAAA GGAAAAAGGC TGTGACATCA TTGTCAAAAA CCACGAATCC 660  
 ATCAGAGTTG GCGGCTACAA TGCCGAAACA GACGAAGTTG AATTCATTTT TGACGACTAC 720  
 TTATATGCTA AGGACGATCT GTTGAAATGG TAA 753

## (2) INFORMATION FOR SEQ ID NO:10:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2520 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: GUT1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TGTATTGGCC ACGATAACCA CCCTTTGTAT ACTGTTTTTG TTTTTCACAT GGTAATAAAC 60  
 GACTTTTATT AAACAACGTA TGTA AAAACA TAACAAGAAT CTACCCATAC AGGCCATTTC 120  
 GTAATTCTTC TCTTCTAATT GGAGTAAAC CATCAATTAA AGGGTGTGGA GTAGCATAGT 180  
 GAGGGGCTGA CTGCATTGAC AAAAAAATTG AAAAAAAAAA AGGAAAAGGA AAGGAAAAAA 240  
 AGACAGCCAA GACTTTTAGA ACGGATAAGG TGTAATAAAA TGTGGGGGGA TGCTTGTCT 300  
 CGAACCATAT AAAATATACC ATGTGGTTTG AGTTGTGGCC GGAACATAC AAATAGTTAT 360  
 ATGTTTCCCT CTCTCTCCG ACTTGTAGTA TTCTCCAAAC GTTACATATT CCGATCAAGC 420  
 CAGCGCCTTT AACTAGTTT AAAACAAGAA CAGAGCCGTA TGTCCAAAAT AATGGAAGAT 480  
 TTACGAAGTG ACTACGTCCC GCTTATCGCC AGTATTGATG TAGGAACGAC CTCATCCAGA 540  
 TGCATTCTGT TCAACAGATG GGGCCAGGAC GTTCAAAAC ACCAAATTGA ATATTCAACT 600  
 TCAGCATCGA AGGGCAAGAT TGGGGTGTCT GGCCTAAGGA GACCTCTAC AGCCCCAGCT 660  
 CGTGAAACAC CAAACGCCGG TGACATCAA ACCAGCGGAA AGCCCATCTT TTCTGCAGAA 720  
 GGCTATGCCA TTCAAGAAAC CAAATTCCTA AAAATCGAGG AATTGGACTT GGAATCCAT 780  
 AACGAACCCA CGTTGAAGTT CCCCAAACCG GGTGGGTTG AGTGCCATCC GCAGAAATTA 840

-- 64 --

CTGGTGAACG TCGTCCAATG CCTTGCCTCA AGTTTGCTCT CTCTGCAGAC TATCAACAGC 900  
GAACGTGTAG CAAACGGTCT CCCACCTTAC AAGGTAATAT GCATGGGTAT AGCAAACATG 960  
AGAGAAACCA CAATTCTGTG GTCCCGCCGC ACAGGAAAAC CAATTGTAA CTACGGTATT 1020  
GTTTGAACG ACACCAGAAC GATCAAAATC GTTAGAGACA AATGGCAAAA CACTAGCGTC 1080  
GATAGGCAAC TGCAGCTTAG ACAGAAGACT GGATTGCCAT TGCTCTCCAC GTATTTCTCC 1140  
TGTTCCAAGC TCGCTGGTT CCTCGACAAT GAGCCTCTGT GTACCAAGGC GTATGAGGAG 1200  
AACGACCTGA TGTTCCGCAC TGTGGACACA TGGCTGATTT ACCAATTAAC TAAACAAAAG 1260  
GCGTTCGTTT CTGACGTAAC CAACGCTTCC AGAACTGGAT TTATGAACCT CTCCACTTTA 1320  
AAGTACGACA ACGAGTTGCT GGAATTTTGG GGTATTGACA AGAACCTGAT TCACATGCCC 1380  
GAAATTGTGT CCTCATCTCA ATACTACGGT GACTTTGGCA TTCCTGATTG GATAATGGAA 1440  
AAGCTACACG ATTCGCCAAA AACAGTACTG CGAGATCTAG TCAAGAGAAA CCTGCCCATA 1500  
CAGGGCTGTC TGGGCGACCA AAGCGCATCC ATGGTGGGGC AACTCGCTTA CAAACCCGGT 1560  
GCTGCAAAAT GTACTTATGG TACCGGTTGC TTTTACTGT ACAATACGGG GACCAAAAAA 1620  
TTGATCTCCC AACATGGCGC ACTGACGACT CTAGCATTTT GGTTCCCACA TTTGCAAGAG 1680  
TACGGTGGCC AAAAACCAGA ATTGAGCAAG CCACATTTTG CATTAGAGGG TTCCGTCGCT 1740  
GTGGCTGGTG CTGTGGTCCA ATGGCTACGT GATAATTTAC GATTGATCGA TAAATCAGAG 1800  
GATGTCGGAC CGATTGCATC TACGGTTCCT GATTCTGGTG GCGTAGTTTT CGTCCCCGCA 1860  
TTTAGTGGCC TATTCGCTCC CTATTGGGAC CCAGATGCCA GAGCCACCAT AATGGGGATG 1920  
TCTCAATTCA CTA CTGCTC CCACATCGCC AGAGCTGCCG TGGAAGGTGT TTGCTTTCAA 1980  
GCCAGGGCTA TCTTGAAGGC AATGAGTTCT GACGCGTTTG GTGAAGGTTC CAAAGACAGG 2040  
GACTTTTTAG AGGAAATTTT CGACGTCACA TATGAAAAGT CGCCCCTGTC GGTTCGGCA 2100  
GTGGATGGCG GGATGTCGAG GTCTAATGAA GTCATGCAAA TTCAAGCCGA TATCCTAGGT 2160  
CCCTGTGTCA AAGTCAGAAG GTCTCCGACA GCGGAATGTA CCGCATTGGG GGCAGCCATT 2220  
GCAGCCAATA TGGCTTTCAA GGATGTGAAC GAGCGCCCAT TATGGAAGGA CCTACACGAT 2280  
GTTAAGAAAT GGGTCTTTTA CAATGGAATG GAGAAAAACG AACAAATATC ACCAGAGGCT 2340  
CATCCAAACC TTAAGATATT CAGAAGTGAA TCCGACGATG CTGAAAGGAG AAAGCATTGG 2400  
AAGTATTGGG AAGTTGCCGT GGAAAGATCC AAAGGTTGGC TGAAGGACAT AGAAGGTGAA 2460  
CACGAACAGG TTCTAGAAAA CTCCAATAA CAACATAAAT AATTTCTATT AACAATGTAA 2520

-- 65 --

## (2) INFORMATION FOR SEQ ID NO:11:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 391 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: protein

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPD1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```

Met Ser Ala Ala Ala Asp Arg Leu Asn Leu Thr Ser Gly His Leu Asn
1           5           10           15

Ala Gly Arg Lys Arg Ser Ser Ser Ser Val Ser Leu Lys Ala Ala Glu
20           25           30

Lys Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr Thr
35           40           45

Ile Ala Lys Val Val Ala Glu Asn Cys Lys Gly Tyr Pro Glu Val Phe
50           55           60

Ala Pro Ile Val Gln Met Trp Val Phe Glu Glu Glu Ile Asn Gly Glu
65           70           75           80

Lys Leu Thr Glu Ile Ile Asn Thr Arg His Gln Asn Val Lys Tyr Leu
85           90           95

Pro Gly Ile Thr Leu Pro Asp Asn Leu Val Ala Asn Pro Asp Leu Ile
100          105          110

Asp Ser Val Lys Asp Val Asp Ile Ile Val Phe Asn Ile Pro His Gln
115          120          125

Phe Leu Pro Arg Ile Cys Ser Gln Leu Lys Gly His Val Asp Ser His
130          135          140

Val Arg Ala Ile Ser Cys Leu Lys Gly Phe Glu Val Gly Ala Lys Gly
145          150          155          160

Val Gln Leu Leu Ser Ser Tyr Ile Thr Glu Glu Leu Gly Ile Gln Cys
165          170          175

Gly Ala Leu Ser Gly Ala Asn Ile Ala Thr Glu Val Ala Gln Glu His
180          185          190

Trp Ser Glu Thr Thr Val Ala Tyr His Ile Pro Lys Asp Phe Arg Gly
195          200          205

Glu Gly Lys Asp Val Asp His Lys Val Leu Lys Ala Leu Phe His Arg

```

-- 66 --

```

      210              215              220
Pro Tyr Phe His Val Ser Val Ile Glu Asp Val Ala Gly Ile Ser Ile
225              230              235              240
Cys Gly Ala Leu Lys Asn Val Val Ala Leu Gly Cys Gly Phe Val Glu
      245              250              255
Gly Leu Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Val Gly
      260              265              270
Leu Gly Glu Ile Ile Arg Phe Gly Gln Met Phe Phe Pro Glu Ser Arg
      275              280              285
Glu Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile Thr
      290              295              300
Thr Cys Ala Gly Gly Arg Asn Val Lys Val Ala Arg Leu Met Ala Thr
305              310              315              320
Ser Gly Lys Asp Ala Trp Glu Cys Glu Lys Glu Leu Leu Asn Gly Gln
      325              330              335
Ser Ala Gln Gly Leu Ile Thr Cys Lys Glu Val His Glu Trp Leu Glu
      340              345              350
Thr Cys Gly Ser Val Glu Asp Phe Pro Leu Phe Glu Ala Val Tyr Gln
      355              360              365
Ile Val Tyr Asn Asn Tyr Pro Met Lys Asn Leu Pro Asp Met Ile Glu
      370              375              380
Glu Leu Asp Leu His Glu Asp
385              390

```

## (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 384 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPD2

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```

Met Thr Ala His Thr Asn Ile Lys Gln His Lys His Cys His Glu Asp
1              5              10              15
His Pro Ile Arg Arg Ser Asp Ser Ala Val Ser Ile Val His Leu Lys
      20              25              30

```



-- 67 --

Arg Ala Pro Phe Lys Val Thr Val Ile Gly Ser Gly Asn Trp Gly Thr  
 35 40 45  
 Thr Ile Ala Lys Val Ile Ala Glu Asn Thr Glu Leu His Ser His Ile  
 50 55 60  
 Phe Glu Pro Glu Val Arg Met Trp Val Phe Asp Glu Lys Ile Gly Asp  
 65 70 75 80  
 Glu Asn Leu Thr Asp Ile Ile Asn Thr Arg His Gln Asn Val Lys Tyr  
 85 90 95  
 Leu Pro Asn Ile Asp Leu Pro His Asn Leu Val Ala Asp Pro Asp Leu  
 100 105 110  
 Leu His Ser Ile Lys Gly Ala Asp Ile Leu Val Phe Asn Ile Pro His  
 115 120 125  
 Gln Phe Leu Pro Asn Ile Val Lys Gln Leu Gln Gly His Val Ala Pro  
 130 135 140  
 His Val Arg Ala Ile Ser Cys Leu Lys Gly Phe Glu Leu Gly Ser Lys  
 145 150 155 160  
 Gly Val Gln Leu Leu Ser Ser Tyr Val Thr Asp Glu Leu Gly Ile Gln  
 165 170 175  
 Cys Gly Ala Leu Ser Gly Ala Asn Leu Ala Pro Glu Val Ala Lys Glu  
 180 185 190  
 His Trp Ser Glu Thr Thr Val Ala Tyr Gln Leu Pro Lys Asp Tyr Gln  
 195 200 205  
 Gly Asp Gly Lys Asp Val Asp His Lys Ile Leu Lys Leu Leu Phe His  
 210 215 220  
 Arg Pro Tyr Phe His Val Asn Val Ile Asp Asp Val Ala Gly Ile Ser  
 225 230 235 240  
 Ile Ala Gly Ala Leu Lys Asn Val Val Ala Leu Ala Cys Gly Phe Val  
 245 250 255  
 Glu Gly Met Gly Trp Gly Asn Asn Ala Ser Ala Ala Ile Gln Arg Leu  
 260 265 270  
 Gly Leu Gly Glu Ile Ile Lys Phe Gly Arg Met Phe Phe Pro Glu Ser  
 275 280 285  
 Lys Val Glu Thr Tyr Tyr Gln Glu Ser Ala Gly Val Ala Asp Leu Ile  
 290 295 300  
 Thr Thr Cys Ser Gly Gly Arg Asn Val Lys Val Ala Thr Tyr Met Ala  
 305 310 315 320  
 Lys Thr Gly Lys Ser Ala Leu Glu Ala Glu Lys Glu Leu Leu Asn Gly  
 325 330 335

-- 68 --

Gln Ser Ala Gln Gly Ile Ile Thr Cys Arg Glu Val His Glu Trp Leu  
                   340                                  345                                  350

Gln Thr Cys Glu Leu Thr Gln Glu Phe Pro Ile Ile Arg Gly Ser Leu  
                   355                                  360                                  365

Pro Asp Ser Leu Gln Gln Arg Pro His Gly Arg Pro Thr Gly Asp Asp  
                   370                                  375                                  380

## (2) INFORMATION FOR SEQ ID NO:13:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 614 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: protein

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: GUT2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Met Thr Arg Ala Thr Trp Cys Asn Ser Pro Pro Pro Leu His Arg Gln  
 1                                  5                                  10                                  15

Val Ser Arg Arg Asp Leu Leu Asp Arg Leu Asp Lys Thr His Gln Phe  
                   20                                  25                                  30

Asp Val Leu Ile Ile Gly Gly Gly Ala Thr Gly Thr Gly Cys Ala Leu  
                   35                                  40                                  45

Asp Ala Ala Thr Arg Gly Leu Asn Val Ala Leu Val Glu Lys Gly Asp  
                   50                                  55                                  60

Phe Ala Ser Gly Thr Ser Ser Lys Ser Thr Lys Met Ile His Gly Gly  
                   65                                  70                                  75                                  80

Val Arg Tyr Leu Glu Lys Ala Phe Trp Glu Phe Ser Lys Ala Gln Leu  
                   85                                  90                                  95

Asp Leu Val Ile Glu Ala Leu Asn Glu Arg Lys His Leu Ile Asn Thr  
                   100                                  105                                  110

Ala Pro His Leu Cys Thr Val Leu Pro Ile Leu Ile Pro Ile Tyr Ser  
                   115                                  120                                  125

Thr Trp Gln Val Pro Tyr Ile Tyr Met Gly Cys Lys Phe Tyr Asp Phe  
                   130                                  135                                  140

Phe Gly Gly Ser Gln Asn Leu Lys Lys Ser Tyr Leu Leu Ser Lys Ser  
                   145                                  150                                  155                                  160

Ala Thr Val Glu Lys Ala Pro Met Leu Thr Thr Asp Asn Leu Lys Ala

-- 69 --

	165		170		175
Ser Leu Val Tyr His Asp Gly Ser Phe Asn Asp Ser Arg Leu Asn Ala	180		185		190
Thr Leu Ala Ile Thr Gly Val Glu Asn Gly Ala Thr Val Leu Ile Tyr	195		200		205
Val Glu Val Gln Lys Leu Ile Lys Asp Pro Thr Ser Gly Lys Val Ile	210		215		220
Gly Ala Glu Ala Arg Asp Val Glu Thr Asn Glu Leu Val Arg Ile Asn	225		230		235
Ala Lys Cys Val Val Asn Ala Thr Gly Pro Tyr Ser Asp Ala Ile Leu	245		250		255
Gln Met Asp Arg Asn Pro Ser Gly Leu Pro Asp Ser Pro Leu Asn Asp	260		265		270
Asn Ser Lys Ile Lys Ser Thr Phe Asn Gln Ile Ser Val Met Asp Pro	275		280		285
Lys Met Val Ile Pro Ser Ile Gly Val His Ile Val Leu Pro Ser Phe	290		295		300
Tyr Ser Pro Lys Asp Met Gly Leu Leu Asp Val Arg Thr Ser Asp Gly	305		310		315
Arg Val Met Phe Phe Leu Pro Trp Gln Gly Lys Val Leu Ala Gly Thr	325		330		335
Thr Asp Ile Pro Leu Lys Gln Val Pro Glu Asn Pro Met Pro Thr Glu	340		345		350
Ala Asp Ile Gln Asp Ile Leu Lys Glu Leu Gln His Tyr Ile Glu Phe	355		360		365
Pro Val Lys Arg Glu Asp Val Leu Ser Ala Trp Ala Gly Val Arg Pro	370		375		380
Leu Val Arg Asp Pro Arg Thr Ile Pro Ala Asp Gly Lys Lys Gly Ser	385		390		395
Ala Thr Gln Gly Val Val Arg Ser His Phe Leu Phe Thr Ser Asp Asn	405		410		415
Gly Leu Ile Thr Ile Ala Gly Gly Lys Trp Thr Thr Tyr Arg Gln Met	420		425		430
Ala Glu Glu Thr Val Asp Lys Val Val Glu Val Gly Gly Phe His Asn	435		440		445
Leu Lys Pro Cys His Thr Arg Asp Ile Lys Leu Ala Gly Ala Glu Glu	450		455		460

-- 70 --

Trp Thr Gln Asn Tyr Val Ala Leu Leu Ala Gln Asn Tyr His Leu Ser  
 465 470 475 480  
 Ser Lys Met Ser Asn Tyr Leu Val Gln Asn Tyr Gly Thr Arg Ser Ser  
 485 490 495  
 Ile Ile Cys Glu Phe Phe Lys Glu Ser Met Glu Asn Lys Leu Pro Leu  
 500 505 510  
 Ser Leu Ala Asp Lys Glu Asn Asn Val Ile Tyr Ser Ser Glu Glu Asn  
 515 520 525  
 Asn Leu Val Asn Phe Asp Thr Phe Arg Tyr Pro Phe Thr Ile Gly Glu  
 530 535 540  
 Leu Lys Tyr Ser Met Gln Tyr Glu Tyr Cys Arg Thr Pro Leu Asp Phe  
 545 550 555 560  
 Leu Leu Arg Arg Thr Arg Phe Ala Phe Leu Asp Ala Lys Glu Ala Leu  
 565 570 575  
 Asn Ala Val His Ala Thr Val Lys Val Met Gly Asp Glu Phe Asn Trp  
 580 585 590  
 Ser Glu Lys Lys Arg Gln Trp Glu Leu Glu Lys Thr Val Asn Phe Ile  
 595 600 605  
 Gln Gly Arg Phe Gly Val  
 610

## (2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 339 amino acids
  - (B) TYPE: amino acid
  - (C) STRANDEDNESS: unknown
  - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:

(A) ORGANISM: GPSA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Met Asn Gln Arg Asn Ala Ser Met Thr Val Ile Gly Ala Gly Ser Tyr  
 1 5 10 15  
 Gly Thr Ala Leu Ala Ile Thr Leu Ala Arg Asn Gly His Glu Val Val  
 20 25 30  
 Leu Trp Gly His Asp Pro Glu His Ile Ala Thr Leu Glu Arg Asp Arg  
 35 40 45  
 Cys Asn Ala Ala Phe Leu Pro Asp Val Pro Phe Pro Asp Thr Leu His  
 50 55 60

-- 71 --

Leu Glu Ser Asp Leu Ala Thr Ala Leu Ala Ala Ser Arg Asn Ile Leu  
 65 70 75 80  
 Val Val Val Pro Ser His Val Phe Gly Glu Val Leu Arg Gln Ile Lys  
 85 90 95  
 Pro Leu Met Arg Pro Asp Ala Arg Leu Val Trp Ala Thr Lys Gly Leu  
 100 105 110  
 Glu Ala Glu Thr Gly Arg Leu Leu Gln Asp Val Ala Arg Glu Ala Leu  
 115 120 125  
 Gly Asp Gln Ile Pro Leu Ala Val Ile Ser Gly Pro Thr Phe Ala Lys  
 130 135 140  
 Glu Leu Ala Ala Gly Leu Pro Thr Ala Ile Ser Leu Ala Ser Thr Asp  
 145 150 155 160  
 Gln Thr Phe Ala Asp Asp Leu Gln Gln Leu Leu His Cys Gly Lys Ser  
 165 170 175  
 Phe Arg Val Tyr Ser Asn Pro Asp Phe Ile Gly Val Gln Leu Gly Gly  
 180 185 190  
 Ala Val Lys Asn Val Ile Ala Ile Gly Ala Gly Met Ser Asp Gly Ile  
 195 200 205  
 Gly Phe Gly Ala Asn Ala Arg Thr Ala Leu Ile Thr Arg Gly Leu Ala  
 210 215 220  
 Glu Met Ser Arg Leu Gly Ala Ala Leu Gly Ala Asp Pro Ala Thr Phe  
 225 230 235 240  
 Met Gly Met Ala Gly Leu Gly Asp Leu Val Leu Thr Cys Thr Asp Asn  
 245 250 255  
 Gln Ser Arg Asn Arg Arg Phe Gly Met Met Leu Gly Gln Gly Met Asp  
 260 265 270  
 Val Gln Ser Ala Gln Glu Lys Ile Gly Gln Val Val Glu Gly Tyr Arg  
 275 280 285  
 Asn Thr Lys Glu Val Arg Glu Leu Ala His Arg Phe Gly Val Glu Met  
 290 295 300  
 Pro Ile Thr Glu Glu Ile Tyr Gln Val Leu Tyr Cys Gly Lys Asn Ala  
 305 310 315 320  
 Arg Glu Ala Ala Leu Thr Leu Leu Gly Arg Ala Arg Lys Asp Glu Arg  
 325 330 335  
 Ser Ser His

(2) INFORMATION FOR SEQ ID NO:15:

-- 72 --

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 501 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: GLPD

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```

Met Glu Thr Lys Asp Leu Ile Val Ile Gly Gly Gly Ile Asn Gly Ala
1           5           10           15

Gly Ile Ala Ala Asp Ala Ala Gly Arg Gly Leu Ser Val Leu Met Leu
          20           25           30

Glu Ala Gln Asp Leu Ala Cys Ala Thr Ser Ser Ala Ser Ser Lys Leu
          35           40           45

Ile His Gly Gly Leu Arg Tyr Leu Glu His Tyr Glu Phe Arg Leu Val
          50           55           60

Ser Glu Ala Leu Ala Glu Arg Glu Val Leu Leu Lys Met Ala Pro His
65           70           75           80

Ile Ala Phe Pro Met Arg Phe Arg Leu Pro His Arg Pro His Leu Arg
          85           90           95

Pro Ala Trp Met Ile Arg Ile Gly Leu Phe Met Tyr Asp His Leu Gly
          100          105          110

Lys Arg Thr Ser Leu Pro Gly Ser Thr Gly Leu Arg Phe Gly Ala Asn
          115          120          125

Ser Val Leu Lys Pro Glu Ile Lys Arg Gly Phe Glu Tyr Ser Asp Cys
          130          135          140

Trp Val Asp Asp Ala Arg Leu Val Leu Ala Asn Ala Gln Met Val Val
          145          150          155          160

Arg Lys Gly Gly Glu Val Leu Thr Arg Thr Arg Ala Thr Ser Ala Arg
          165          170          175

Arg Glu Asn Gly Leu Trp Ile Val Glu Ala Glu Asp Ile Asp Thr Gly
          180          185          190

Lys Lys Tyr Ser Trp Gln Ala Arg Gly Leu Val Asn Ala Thr Gly Pro
          195          200          205

Trp Val Lys Gln Phe Phe Asp Asp Gly Met His Leu Pro Ser Pro Tyr
          210          215          220

```

-- 73 --

Gly Ile Arg Leu Ile Lys Gly Ser His Ile Val Val Pro Arg Val His  
 225 230 235 240  
 Thr Gln Lys Gln Ala Tyr Ile Leu Gln Asn Glu Asp Lys Arg Ile Val  
 245 250 255  
 Phe Val Ile Pro Trp Met Asp Glu Phe Ser Ile Ile Gly Thr Thr Asp  
 260 265 270  
 Val Glu Tyr Lys Gly Asp Pro Lys Ala Val Lys Ile Glu Glu Ser Glu  
 275 280 285  
 Ile Asn Tyr Leu Leu Asn Val Tyr Asn Thr His Phe Lys Lys Gln Leu  
 290 295 300  
 Ser Arg Asp Asp Ile Val Trp Thr Tyr Ser Gly Val Arg Pro Leu Cys  
 305 310 315 320  
 Asp Asp Glu Ser Asp Ser Pro Gln Ala Ile Thr Arg Asp Tyr Thr Leu  
 325 330 335  
 Asp Ile His Asp Glu Asn Gly Lys Ala Pro Leu Leu Ser Val Phe Gly  
 340 345 350  
 Gly Lys Leu Thr Thr Tyr Arg Lys Leu Ala Glu His Ala Leu Glu Lys  
 355 360 365  
 Leu Thr Pro Tyr Tyr Gln Gly Ile Gly Pro Ala Trp Thr Lys Glu Ser  
 370 375 380  
 Val Leu Pro Gly Gly Ala Ile Glu Gly Asp Arg Asp Asp Tyr Ala Ala  
 385 390 395 400  
 Arg Leu Arg Arg Arg Tyr Pro Phe Leu Thr Glu Ser Leu Ala Arg His  
 405 410 415  
 Tyr Ala Arg Thr Tyr Gly Ser Asn Ser Glu Leu Leu Leu Gly Asn Ala  
 420 425 430  
 Gly Thr Val Ser Asp Leu Gly Glu Asp Phe Gly His Glu Phe Tyr Glu  
 435 440 445  
 Ala Glu Leu Lys Tyr Leu Val Asp His Glu Trp Val Arg Arg Ala Asp  
 450 455 460  
 Asp Ala Leu Trp Arg Arg Thr Lys Gln Gly Met Trp Leu Asn Ala Asp  
 465 470 475 480  
 Gln Gln Ser Arg Val Ser Gln Trp Leu Val Glu Tyr Thr Gln Gln Arg  
 485 490 495  
 Leu Ser Leu Ala Ser  
 500

(2) INFORMATION FOR SEQ ID NO:16:

-- 74 --

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 542 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: GLPABC

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

```

Met Lys Thr Arg Asp Ser Gln Ser Ser Asp Val Ile Ile Ile Gly Gly
1           5           10           15

Gly Ala Thr Gly Ala Gly Ile Ala Arg Asp Cys Ala Leu Arg Gly Leu
          20           25           30

Arg Val Ile Leu Val Glu Arg His Asp Ile Ala Thr Gly Ala Thr Gly
          35           40           45

Arg Asn His Gly Leu Leu His Ser Gly Ala Arg Tyr Ala Val Thr Asp
          50           55           60

Ala Glu Ser Ala Arg Glu Cys Ile Ser Glu Asn Gln Ile Leu Lys Arg
          65           70           75           80

Ile Ala Arg His Cys Val Glu Pro Thr Asn Gly Leu Phe Ile Thr Leu
          85           90           95

Pro Glu Asp Asp Leu Ser Phe Gln Ala Thr Phe Ile Arg Ala Cys Glu
          100          105          110

Glu Ala Gly Ile Ser Ala Glu Ala Ile Asp Pro Gln Gln Ala Arg Ile
          115          120          125

Ile Glu Pro Ala Val Asn Pro Ala Leu Ile Gly Ala Val Lys Val Pro
          130          135          140

Asp Gly Thr Val Asp Pro Phe Arg Leu Thr Ala Ala Asn Met Leu Asp
          145          150          155          160

Ala Lys Glu His Gly Ala Val Ile Leu Thr Ala His Glu Val Thr Gly
          165          170          175

Leu Ile Arg Glu Gly Ala Thr Val Cys Gly Val Arg Val Arg Asn His
          180          185          190

Leu Thr Gly Glu Thr Gln Ala Leu His Ala Pro Val Val Val Asn Ala
          195          200          205

Ala Gly Ile Trp Gly Gln His Ile Ala Glu Tyr Ala Asp Leu Arg Ile
          210          215          220

Arg Met Phe Pro Ala Lys Gly Ser Leu Leu Ile Met Asp His Arg Ile

```



-- 75 --

225		230		235		240
Asn Gln His Val Ile Asn Arg Cys Arg Lys Pro Ser Asp Ala Asp Ile						
	245			250		255
Leu Val Pro Gly Asp Thr Ile Ser Leu Ile Gly Thr Thr Ser Leu Arg						
	260		265			270
Ile Asp Tyr Asn Glu Ile Asp Asp Asn Arg Val Thr Ala Glu Glu Val						
	275		280			285
Asp Ile Leu Leu Arg Glu Gly Glu Lys Leu Ala Pro Val Met Ala Lys						
	290		295		300	
Thr Arg Ile Leu Arg Ala Tyr Ser Gly Val Arg Pro Leu Val Ala Ser						
	305	310		315		320
Asp Asp Asp Pro Ser Gly Arg Asn Leu Ser Arg Gly Ile Val Leu Leu						
	325		330			335
Asp His Ala Glu Arg Asp Gly Leu Asp Gly Phe Ile Thr Ile Thr Gly						
	340		345			350
Gly Lys Leu Met Thr Tyr Arg Leu Met Ala Glu Trp Ala Thr Asp Ala						
	355		360			365
Val Cys Arg Lys Leu Gly Asn Thr Arg Pro Cys Thr Thr Ala Asp Leu						
	370		375			380
Ala Leu Pro Gly Ser Gln Glu Pro Ala Glu Val Thr Leu Arg Lys Val						
	385		390		395	400
Ile Ser Leu Pro Ala Pro Leu Arg Gly Ser Ala Val Tyr Arg His Gly						
	405		410			415
Asp Arg Thr Pro Ala Trp Leu Ser Glu Gly Arg Leu His Arg Ser Leu						
	420		425			430
Val Cys Glu Cys Glu Ala Val Thr Ala Gly Glu Val Gln Tyr Ala Val						
	435		440			445
Glu Asn Leu Asn Val Asn Ser Leu Leu Asp Leu Arg Arg Arg Thr Arg						
	450		455			460
Val Gly Met Gly Thr Cys Gln Gly Glu Leu Cys Ala Cys Arg Ala Ala						
	465		470		475	480
Gly Leu Leu Gln Arg Phe Asn Val Thr Thr Ser Ala Gln Ser Ile Glu						
	485		490			495
Gln Leu Ser Thr Phe Leu Asn Glu Arg Trp Lys Gly Val Gln Pro Ile						
	500		505			510
Ala Trp Gly Asp Ala Leu Arg Glu Ser Glu Phe Thr Arg Trp Val Tyr						
	515		520			525

-- 76 --

Gln Gly Leu Cys Gly Leu Glu Lys Glu Gln Lys Asp Ala Leu  
 530 535 540

## (2) INFORMATION FOR SEQ ID NO:17:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 250 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: protein

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: GPP2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met Gly Leu Thr Thr Lys Pro Leu Ser Leu Lys Val Asn Ala Ala Leu  
 1 5 10 15

Phe Asp Val Asp Gly Thr Ile Ile Ile Ser Gln Pro Ala Ile Ala Ala  
 20 25 30

Phe Trp Arg Asp Phe Gly Lys Asp Lys Pro Tyr Phe Asp Ala Glu His  
 35 40 45

Val Ile Gln Val Ser His Gly Trp Arg Thr Phe Asp Ala Ile Ala Lys  
 50 55 60

Phe Ala Pro Asp Phe Ala Asn Glu Glu Tyr Val Asn Lys Leu Glu Ala  
 65 70 75 80

Glu Ile Pro Val Lys Tyr Gly Glu Lys Ser Ile Glu Val Pro Gly Ala  
 85 90 95

Val Lys Leu Cys Asn Ala Leu Asn Ala Leu Pro Lys Glu Lys Trp Ala  
 100 105 110

Val Ala Thr Ser Gly Thr Arg Asp Met Ala Gln Lys Trp Phe Glu His  
 115 120 125

Leu Gly Ile Arg Arg Pro Lys Tyr Phe Ile Thr Ala Asn Asp Val Lys  
 130 135 140

Gln Gly Lys Pro His Pro Glu Pro Tyr Leu Lys Gly Arg Asn Gly Leu  
 145 150 155 160

Gly Tyr Pro Ile Asn Glu Gln Asp Pro Ser Lys Ser Lys Val Val Val  
 165 170 175

Phe Glu Asp Ala Pro Ala Gly Ile Ala Ala Gly Lys Ala Ala Gly Cys  
 180 185 190

Lys Ile Ile Gly Ile Ala Thr Thr Phe Asp Leu Asp Phe Leu Lys Glu  
 195 200 205

-- 77 --

Lys Gly Cys Asp Ile Ile Val Lys Asn His Glu Ser Ile Arg Val Gly  
 210 215 220

Gly Tyr Asn Ala Glu Thr Asp Glu Val Glu Phe Ile Phe Asp Asp Tyr  
 225 230 235 240

Leu Tyr Ala Lys Asp Asp Leu Leu Lys Trp  
 245 250

## (2) INFORMATION FOR SEQ ID NO:18:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 709 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: protein

## (vi) ORIGINAL SOURCE:

- (A) ORGANISM: GUT1

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Met Phe Pro Ser Leu Phe Arg Leu Val Val Phe Ser Lys Arg Tyr Ile  
 1 5 10 15

Phe Arg Ser Ser Gln Arg Leu Tyr Thr Ser Leu Lys Gln Glu Gln Ser  
 20 25 30

Arg Met Ser Lys Ile Met Glu Asp Leu Arg Ser Asp Tyr Val Pro Leu  
 35 40 45

Ile Ala Ser Ile Asp Val Gly Thr Thr Ser Ser Arg Cys Ile Leu Phe  
 50 55 60

Asn Arg Trp Gly Gln Asp Val Ser Lys His Gln Ile Glu Tyr Ser Thr  
 65 70 75 80

Ser Ala Ser Lys Gly Lys Ile Gly Val Ser Gly Leu Arg Arg Pro Ser  
 85 90 95

Thr Ala Pro Ala Arg Glu Thr Pro Asn Ala Gly Asp Ile Lys Thr Ser  
 100 105 110

Gly Lys Pro Ile Phe Ser Ala Glu Gly Tyr Ala Ile Gln Glu Thr Lys  
 115 120 125

Phe Leu Lys Ile Glu Glu Leu Asp Leu Asp Phe His Asn Glu Pro Thr  
 130 135 140

Leu Lys Phe Pro Lys Pro Gly Trp Val Glu Cys His Pro Gln Lys Leu  
 145 150 155 160

Leu Val Asn Val Val Gln Cys Leu Ala Ser Ser Leu Leu Ser Leu Gln

-- 78 --

	165		170		175
Thr Ile Asn Ser Glu Arg Val Ala Asn Gly Leu Pro Pro Tyr Lys Val	180		185		190
Ile Cys Met Gly Ile Ala Asn Met Arg Glu Thr Thr Ile Leu Trp Ser	195		200		205
Arg Arg Thr Gly Lys Pro Ile Val Asn Tyr Gly Ile Val Trp Asn Asp	210		215		220
Thr Arg Thr Ile Lys Ile Val Arg Asp Lys Trp Gln Asn Thr Ser Val	225		230		240
Asp Arg Gln Leu Gln Leu Arg Gln Lys Thr Gly Leu Pro Leu Leu Ser	245		250		255
Thr Tyr Phe Ser Cys Ser Lys Leu Arg Trp Phe Leu Asp Asn Glu Pro	260		265		270
Leu Cys Thr Lys Ala Tyr Glu Glu Asn Asp Leu Met Phe Gly Thr Val	275		280		285
Asp Thr Trp Leu Ile Tyr Gln Leu Thr Lys Gln Lys Ala Phe Val Ser	290		295		300
Asp Val Thr Asn Ala Ser Arg Thr Gly Phe Met Asn Leu Ser Thr Leu	305		310		320
Lys Tyr Asp Asn Glu Leu Leu Glu Phe Trp Gly Ile Asp Lys Asn Leu	325		330		335
Ile His Met Pro Glu Ile Val Ser Ser Ser Gln Tyr Tyr Gly Asp Phe	340		345		350
Gly Ile Pro Asp Trp Ile Met Glu Lys Leu His Asp Ser Pro Lys Thr	355		360		365
Val Leu Arg Asp Leu Val Lys Arg Asn Leu Pro Ile Gln Gly Cys Leu	370		375		380
Gly Asp Gln Ser Ala Ser Met Val Gly Gln Leu Ala Tyr Lys Pro Gly	385		390		400
Ala Ala Lys Cys Thr Tyr Gly Thr Gly Cys Phe Leu Leu Tyr Asn Thr	405		410		415
Gly Thr Lys Lys Leu Ile Ser Gln His Gly Ala Leu Thr Thr Leu Ala	420		425		430
Phe Trp Phe Pro His Leu Gln Glu Tyr Gly Gly Gln Lys Pro Glu Leu	435		440		445
Ser Lys Pro His Phe Ala Leu Glu Gly Ser Val Ala Val Ala Gly Ala	450		455		460

-- 79 --

Val Val Gln Trp Leu Arg Asp Asn Leu Arg Leu Ile Asp Lys Ser Glu  
 465 470 475 480  
 Asp Val Gly Pro Ile Ala Ser Thr Val Pro Asp Ser Gly Gly Val Val  
 485 490 495  
 Phe Val Pro Ala Phe Ser Gly Leu Phe Ala Pro Tyr Trp Asp Pro Asp  
 500 505 510  
 Ala Arg Ala Thr Ile Met Gly Met Ser Gln Phe Thr Thr Ala Ser His  
 515 520 525  
 Ile Ala Arg Ala Ala Val Glu Gly Val Cys Phe Gln Ala Arg Ala Ile  
 530 535 540  
 Leu Lys Ala Met Ser Ser Asp Ala Phe Gly Glu Gly Ser Lys Asp Arg  
 545 550 555 560  
 Asp Phe Leu Glu Glu Ile Ser Asp Val Thr Tyr Glu Lys Ser Pro Leu  
 565 570 575  
 Ser Val Leu Ala Val Asp Gly Gly Met Ser Arg Ser Asn Glu Val Met  
 580 585 590  
 Gln Ile Gln Ala Asp Ile Leu Gly Pro Cys Val Lys Val Arg Arg Ser  
 595 600 605  
 Pro Thr Ala Glu Cys Thr Ala Leu Gly Ala Ala Ile Ala Ala Asn Met  
 610 615 620  
 Ala Phe Lys Asp Val Asn Glu Arg Pro Leu Trp Lys Asp Leu His Asp  
 625 630 635 640  
 Val Lys Lys Trp Val Phe Tyr Asn Gly Met Glu Lys Asn Glu Gln Ile  
 645 650 655  
 Ser Pro Glu Ala His Pro Asn Leu Lys Ile Phe Arg Ser Glu Ser Asp  
 660 665 670  
 Asp Ala Glu Arg Arg Lys His Trp Lys Tyr Trp Glu Val Ala Val Glu  
 675 680 685  
 Arg Ser Lys Gly Trp Leu Lys Asp Ile Glu Gly Glu His Glu Gln Val  
 690 695 700  
 Leu Glu Asn Phe Gln  
 705

## (2) INFORMATION FOR SEQ ID NO:19:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12145 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

-- 80 --

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: PHK28-26

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GTCGACCACC	ACGGTGGTGA	CTTTAATGCC	GCTCTCATGC	AGCAGCTCGG	TGGCGGTCTC	60
AAAATTCAGG	ATGTCGCCGG	TATAGTTTTT	GATAATCAGC	AAGACGCCTT	CGCCGCCGTC	120
AATTTGCATC	GCGCATTCAA	ACATTTTGTC	CGGCGTCGGC	GAGGTGAATA	TTTCCCCCGG	180
ACAGGCGCCG	GAGAGCATGC	CCTGGCCGAT	ATAGCCGCAG	TGCATCGGTT	CATGTCCGCT	240
GCCGCCGCCG	GAGAGCAGGG	CCACCTTGCC	AGCCACCGGC	GCGTCGGTGC	GGGTCACATA	300
CAGCGGGTCC	TGATGCAGGG	TCAGCTGCGG	ATGGGCTTTA	GCCAGCCCCT	GTAATTGTTC	360
ATTCAGTACA	TCTTCAACAC	GGTTAATCAG	CTTTTTCATT	ATTCAGTGCT	CCGTTGGAGA	420
AGGTTTCGATG	CCGCCTCTCT	GCTGGCGGAG	GCGGTCATCG	CGTAGGGGTA	TCGTCTGACG	480
GTGGAGCGTG	CCTGGCGATA	TGATGATTCT	GGCTGAGCGG	ACGAAAAAAA	GAATGCCCCG	540
ACGATCGGGT	TTCATTACGA	AACATTGCTT	CCTGATTTTG	TTTCTTTATG	GAACGTTTTT	600
GCTGAGGATA	TGGTGAAAAT	GCGAGCTGGC	GCGCTTTTTT	TCTTCTGCCA	TAAGCGGCGG	660
TCAGGATAGC	CGGCGAAGCG	GGTGGGAAAA	AATTTTTTGC	TGATTTTCTG	CCGACTGCGG	720
GAGAAAAGGC	GGTCAAACAC	GGAGGATTGT	AAGGGCATT	TGCGGCAAAG	GAGCGGATCG	780
GGATCGCAAT	CCTGACAGAG	ACTAGGGTTT	TTTGTTCCAA	TATGGAACGT	AAAAAATTAA	840
CCTGTGTTTC	ATATCAGAAC	AAAAAGGCGA	AAGATTTTTT	TGTTCCCTGC	CGGCCCTACA	900
GTGATCGCAC	TGCTCCGGTA	CGCTCCGTTC	AGGCCGCGCT	TCACTGGCCG	GCGCGGATAA	960
CGCCAGGGCT	CATCATGTCT	ACATGCGCAC	TTATTTGAGG	GTGAAAGGAA	TGCTAAAAGT	1020
TATTCAATCT	CCAGCCAAAT	ATCTTCAGGG	TCCTGATGCT	GCTGTTCTGT	TCGGTCAATA	1080
TGCCAAAAAC	CTGGCGGAGA	GCTTCTTCGT	CATCGCTGAC	GATTTTCGTAA	TGAAGCTGGC	1140
GGGAGAGAAA	GTGGTGAATG	GCCTGCAGAG	CCACGATATT	CGCTGCCATG	CGGAACGGTT	1200
TAACGGCGAA	TGCAGCCATG	CGGAAATCAA	CCGTCTGATG	GCGATTTTGC	AAAAACAGGG	1260
CTGCCGCGGC	GTGGTCGGGA	TCGGCGGTGG	TAAAACCCTC	GATACCGCGA	AGGCGATCGG	1320
TTACTACCAG	AAGCTGCCGG	TGGTGGTGAT	CCCGACCATC	GCCTCGACCG	ATGCGCCAAC	1380
CAGCGCGCTG	TCGGTGATCT	ACACCGAAGC	GGGCGAGTTT	GAAGAGTATC	TGATCTATCC	1440
GAAAAACCCG	GATATGGTGG	TGATGGACAC	GGCGATTATC	GCCAAAGCGC	CGGTACGCCT	1500

-- 81 --

GCTGGTCTCC	GGCATGGGCG	ATGCGCTCTC	CACCTGGTTC	GAGGCCAAAG	CTTGCTACGA	1560
TGCGCGCGCC	ACCAGCATGG	CCGGAGGACA	GTCCACCGAG	GCGGCGCTGA	GCCTCGCCCG	1620
CCTGTGCTAT	GATACGCTGC	TGGCGGAGGG	CGAAAAGGCC	CGTCTGGCGG	CGCAGGCCCG	1680
GGTAGTGACC	GAAGCGCTGG	AGCGCATCAT	CGAGGCGAAC	ACTTACCTCA	GCGGCATTGG	1740
CTTTGAAAGC	AGTGGCCTGG	CCGCTGCCCC	TGCAATCCAC	AACGGTTTCA	CCATTCTTGA	1800
AGAGTGCCAT	CACCTGTATC	ACGGTGAGAA	AGTGGCCTTC	GGTACCCTGG	CGCAGCTGGT	1860
GCTGCAGAAC	AGCCCGATGG	ACGAGATTGA	AACGGTGCAG	GGCTTCTGCC	AGCGCGTCGG	1920
CCTGCCGGTG	ACGCTCGCGC	AGATGGGCGT	CAAAGAGGGG	ATCGACGAGA	AAATCGCCGC	1980
GGTGGCGAAA	GCTACCTGCG	GGGAAGGGGA	AACCATCCAT	AATATGCCGT	TTGCGGTGAC	2040
CCCGGAGAGC	GTCCATGCCG	CTATCCTCAC	CGCCGATCTG	TTAGGCCAGC	AGTGGCTGGC	2100
GCGTTAATTC	GCGGTGGCTA	AACCGCTGGC	CCAGGTCAGC	GGTTTTTCTT	TCTCCCCTCC	2160
GGCAGTCGCT	GCCGGAGGGG	TTCTCTATGG	TACAACGCGG	AAAAGGATAT	GACTGTTCAG	2220
ACTCAGGATA	CCGGGAAGGC	GGTCTCTTCC	GTCATTGCCC	AGTCATGGCA	CCGCTGCAGC	2280
AAGTTTATGC	AGCGCGAAAC	CTGGCAAACG	CCGCACCAGG	CCCAGGGCCT	GACCTTCGAC	2340
TCCATCTGTC	GGCGTAAAC	CGCGCTGCTC	ACCATCGGCC	AGGCGGCGCT	GGAAGACGCC	2400
TGGGAGTTTA	TGGACGGCCG	CCCCTGCGCG	CTGTTTATTC	TTGATGAGTC	CGCCTGCATC	2460
CTGAGCCGTT	GCGGCGAGCC	GCAAACCCTG	GCCCAGCTGG	CTGCCCTGGG	ATTCGCGAC	2520
GGCAGCTATT	GTGCGGAGAG	CATTATCGGC	ACCTGCGCGC	TGTCGCTGGC	CGCGATGCAG	2580
GGCCAGCCGA	TCAACACCGC	CGGCGATCGG	CATTTTAAGC	AGGCGCTACA	GCCATGGAGT	2640
TTTGTGCTCGA	CGCCGGTGTT	TGATAACCAC	GGGCGGCTGT	TCGGCTCTAT	CTCGCTTTGC	2700
TGTCTGGTCG	AGCACCAGTC	QAGCGCCGAC	CTCTCCCTGA	CGCTGGCCAT	CGCCCGCGAG	2760
GTGGGTAACT	CCCTGCTTAC	CGACAGCCTG	CTGGCGGAAT	CCAACCGTCA	CCTCAATCAG	2820
ATGTACGGCC	TGCTGGAGAG	CATGGACGAT	GGGGTGATGG	CGTGGAACGA	ACAGGGCGTG	2880
CTGCAGTTTC	TCAATGTTCA	GGCGGCGAGA	CTGCTGCATC	TTGATGCTCA	GGCCAGCCAG	2940
GGGAAAAATA	TCGCCGATCT	GGTGACCCTC	CCGGCGCTGC	TGCGCCGCGC	CATCAAACAC	3000
GCCCGCGGCC	TGAATCACGT	CGAAGTCACC	TTTGAAAGTC	AGCATCAGTT	TGTCGATGCG	3060
GTGATCACCT	TAAAACCGAT	TGTCGAGGCG	CAAGGCAACA	GTTTTATTCT	GCTGCTGCAT	3120
CCGGTGGAGC	AGATGCGGCA	GCTGATGACC	AGCCAGCTCG	GTAAAGTCAG	CCACACCTTT	3180

-- 82 --

GAGCAGATGT CTGCCGACGA TCCGGAAACC CGACGCCTGA TCCACTTTGG CCGCCAGGCG	3240
GCGCGCGGCG GCTTCCCGGT GCTACTGTGC GGC GAAGAGG GGGTCGGGAA AGAGCTGCTG	3300
AGCCAGGCTA TTCACAATGA AAGCGAACGG GCGGGCGGCC CCTACATCTC CGTCAACTGC	3360
CAGCTATATG CCGACAGCGT GCTGGGCCAG GACTTTATGG GCAGCGCCCC TACCGACGAT	3420
GAAATGGTC GCCTGAGCCG CTTGAGCTG GCCAACGGCG GCACCCTGTT TCTGAAAAG	3480
ATCGAGTATC TGGCGCCGGA GCTGCAGTCG GCTCTGCTGC AGGTGATTAA GCAGGGCGTG	3540
CTCACC CGC TCGACGCCG GCGCCTGATC CCGGTGGATG TGAAGGTGAT TGCCACCACC	3600
ACCGTCGATC TGGCCAATCT GGTGGAACAG AACCGCTTTA GCCGCCAGCT GTACTATGCG	3660
CTGCACTCCT TTGAGATCGT CATCCCGCCG CTGCGCGCCC GACGCAACAG TATTCCGTCG	3720
CTGGTGCATA ACCGGTTGAA GAGCCTGGAG AAGCGTTTCT CTTCGCGACT GAAAGTGGAC	3780
GATGACGCGC TGGCACAGCT GGTGGCCTAC TCGTGGCCGG GGAATGATTT TGAGCTCAAC	3840
AGCGTCATTG AGAATATCGC CATCAGCAGC GACAACGGCC ACATTCGCCT GAGTAATCTG	3900
CCGGAATATC TCTTTTCCGA GCGGCCGGGC GGGGATAGCG CGTCATCGCT GCTGCCGGCC	3960
AGCCTGACTT TTAGCGCCAT CGAAAAGGAA GCTATTATTC ACGCCGCCCG GGTGACCAGC	4020
GGGCGGGTGC AGGAGATGTC GCAGCTGCTC AATATCGGCC GCACCACCCT GTGGCGCAA	4080
ATGAAGCAGT ACGATATTGA CGCCAGCCAG TTCAAGCGCA AGCATCAGGC CTAGTCTCTT	4140
CGATTCGCGC CATGGAGAAC AGGGCATCCG ACAGGCGATT GCTGTAGCGT TTGAGCGCGT	4200
CGCGCAGCGG ATGCGCGCGG TCCATGGCCG TCAGCAGGCG TTCGAGCCGA CGGGACTGGG	4260
TGCGCGCCAC GTGCAGCTGG GCAGAGGCGA GATTCCTCCC CGGGATCACG AACTGTTTTA	4320
ACGGGCCGCT CTCGGCCATA TTGCGGTCGA TAAGCCGCTC CAGGGCGGTG ATCTCCTCTT	4380
CGCCGATCGT CTGGCTCAGG CGGGTCAGGC CCCGCGCATC GCTGGCCAGT TCAGCCCCCA	4440
GCACGAACAG CGTCTGCTGA ATATGGTGCA GGCTTTCCCG CAGCCCGGCG TCGCGGGTCG	4500
TGGCGTAGCA GACGCCCAGC TGGGATATCA GTTCATCGAC GGTGCCGTAG GCCTCGACGC	4560
GAATATGGTC TTTCTCGATG CGGCTGCCGC CGTACAGGGC GGTGGTGCCT TTATCCCCGG	4620
TGCGGGTATA GATACGATAC ATTCAGTTTC TCTCACTTAA CGGCAGGACT TTAACCAGCT	4680
GCCCCGGCGTT GCGCCGAGC GTACGCAGTT GATCGTCGCT ATCGGTGACG TGTCCGGTAG	4740
CCAGCGGCGC GTCCGCCGGC AGCTGGGCAT GAGTGAGGGC TATCTCGCCG GACGCGCTGA	4800
GCCCGATACC CACCCGCAGG GCGGAGCTTC TGGCCGCCAG GCGCCCGAGC GCAGCGGCGT	4860



-- 83 --

CACCGCCTCC	GTCATAGGTT	ATGGTCTGGC	AGGGGACCCC	CTGCTCCTCC	AGCCCCCAGC	4920
ACAGCTCATT	GATGGCGCCG	GCATGGTGCC	CGCGCGGATC	GTAAACAGG	CGTACGCCTG	4980
CGGGTGAAAG	CGACATGACG	GTCCCCTCGT	TAACACTCAG	AATGCCTGGC	GGAAAATCGC	5040
GGCAATCTCC	TGCTCGTTGC	CTTTACGCGG	GTTCGAGAAC	GCATTGCCGT	CTTTTAGAGC	5100
CATCTCCGCC	ATGTAGGGGA	AGTCGGCCTC	TTTTACCCCC	AGATCGCGCA	GATGCTGCGG	5160
AATACCGATA	TCCATCGACA	GACGCGTGAT	AGCGGCGATG	GCTTTTTCCTG	CCGCGTCGAG	5220
AGTGGACAGT	CCGGTGATAT	TTTCGCCCCAT	CAGTTCAGCG	ATATCGGCGA	ATTTCTCCGG	5280
GTTGGCGATC	AGGTTGTAGC	GCGCCACATG	CGGCAGCAGG	ACAGCGTTGG	CCACGCCGTG	5340
CGGCATGTCC	TACAGGCCGC	CCAGCTGGTG	CGCCATGGCG	TGCACGTAGC	CGAGGTTGGC	5400
GTTATTGAAA	GCCATCCCGG	CCAGCAGAGA	AGCATAGGCC	ATGTTTTCCTG	GCGCCTGCAG	5460
ATTGCTGCCG	AGGGCCACGG	CCTGGCGCAG	GTTGCGGGCG	ATGAGGCGGA	TCGCCTGCAT	5520
GGCGGCGGCG	TCCGTCACCG	GGTTAGCGTC	TTTGAGAGATA	TAGGCCTCTA	CGGCGTGGGT	5580
CAGGGCATCC	ATCCCGGTCG	C CGCGGTCAG	GGCGGCCGGT	TTACCGATCA	TCAGCAGTGG	5640
ATCGTTGATA	GAGACCGACG	GCAGTTTGCG	CCAGCTGACG	ATCACAAACT	TCACTTTGGT	5700
TTCGGTGTTG	GTCAGGACGC	AGTGGCGGGT	GACCTCGCTG	GCGGTGCCGG	CGGTGGTATT	5760
GACCGCGACG	ATAGGCGGCA	GCGGGTTGGT	CAGGGTCTCG	ATTCCGGCAT	ACTGGTACAG	5820
ATCGCCCTCA	TGGGTGGCGG	CGATGCCGAT	GCCTTTGCCG	CAATCGTGCG	GGCTGCCGCC	5880
GCCCACGGTG	ACGATGATGT	CGCACTGTTC	GCGGCGAAAC	ACGGCGAGGC	CGTCGCGCAC	5940
GTTGGTGTCT	TTCGGGTTCG	GCTCGACGCC	GTCAAAGATC	GCCACCTCGA	TCCCGGCCCTC	6000
CCGCAGATAA	TGCAGGGTTT	TGTCCACCGC	GCCATCTTTA	ATTGCCCGCA	GGCCTTTGTC	6060
GGTGACCAGC	AGGGCTTTTT	TCCCCCCCAG	CAGCTGGCAG	CGTTCGCCGA	CTACGGAAAT	6120
GGCGTTGGGG	CCAAAAAAGT	TAACGTTTGG	CACCAGATAA	TCAAACATAC	GATAGCTCAT	6180
AATATACCTT	CTCGCTTCAG	GTTATAATGC	GGAAAAACAA	TCCAGGGCGC	ACTGGGCTAA	6240
TAATTGATCC	TGCTCGACCG	TACCGCCGCT	AACGCCGACG	GCGCCAATTA	CCTGCTCATT	6300
AAAAATAACT	GGCAGGCCGC	CGCCAAAAAT	AATAATTTCG	TGTTGGTTGG	TTAGCTGCAG	6360
ACCGTACAGA	GATTGTCCTG	GCTGGACCGC	TGACGTAATT	TCATGGGTAC	CTTGCTTCAG	6420
GCTGCAGGCG	CTCCAGGCTT	TATTCAGGGA	AATATCGCAG	CTGGAGACGA	AGGCCTCGTC	6480
CATCCGCTGG	ATAAGCAGCG	TGTTGCCTCC	GCGGTCAACT	ACGGAACA	CCACCGCCAC	6540

-- 84 --

GTTGATCTCA GTGGCTTTTT TTTCCACCGC CGCCGCCATT TGCTGGGCGG CGGCCAGGGT 6600  
GATTGTCTGA ACTTGTTGGC TCTTGTTTCAT CATTCTCTCC CGCACCAGGA TAACGCTGGC 6660  
GCGAATAGTC AGTAGGGGGC GATAGTAAAA AACTATTACC ATTCGGTTGG CTGCTTTAT 6720  
TTTTGTCAGC GTTATTTTGT CGCCCGCCAT GATTTAGTCA ATAGGGTTAA AATAGCGTCG 6780  
GAAAAACGTA ATTAAGGGCG TTTTTTATTA ATTGATTTAT ATCATTGCGG GCGATCACAT 6840  
TTTTTATTTT TGCCGCCGGA GTAAAGTTTC ATAGTGAAAC TGTCGGTAGA TTTCGTGTGC 6900  
CAAATTGAAA CGAAATTAAA TTTATTTTTT TCACCACTGG CTCATTTAAA GTTCCGCTAT 6960  
TGCCGGTAAT GGCCGGGCGG CAACGACGCT GGCCCGGCGT ATTCGCTACC GTCTGCGGAT 7020  
TTCACCTTTT GAGCCGATGA AGAATGAAAA GATCAAAACG ATTTGCAGTA CTGGCCCAGC 7080  
GCCCCGTCAA TCAGGACGGG CTGATTGGCG AGTGGCCTGA AGAGGGGCTG ATCGCCATGG 7140  
ACAGCCCCTT TGACCCGGTC TCTTCAGTAA AAGTGGACAA CGGTCTGATC GTCGAACTGG 7200  
ACGGCAAACG CCGGGACCAG TTTGACATGA TCGACCGATT TATCGCCGAT TACGCGATCA 7260  
ACGTTGAGCG CACAGAGCAG GCAATGCGCC TGGAGGCGGT GGAAATAGCC CGTATGCTGG 7320  
TGGATATTCA CGTCAGCCGG GAGGAGATCA TTGCCATCAC TACCGCCATC ACGCCGGCCA 7380  
AAGCGGTCGA GGTGATGGCG CAGATGAACG TGGTGGAGAT GATGATGGCG CTGCAGAAGA 7440  
TGCGTGCCCG CCGGACCCCC TCCAACCACT GCCACGTCAC CAATCTCAA GATAATCCGG 7500  
TGCAGATTGC CGCTGACGCC GCCGAGGCCG GGATCCGCGG CTTCTCAGAA CAGGAGACCA 7560  
CGGTGCGTAT CGCGCGCTAC GCGCCGTTTA ACGCCCTGGC GCTGTTGGTC GGTTCGCAGT 7620  
GCGGCCGCCC CGGCGTGTTG ACGCAGTGCT CGGTGGAAGA GGCCACCGAG CTGGAGCTGG 7680  
GCATGCGTGG CTTAACCAGC TACGCCGAGA CGGTGTCGGT CTACGGCACC GAAGCGGTAT 7740  
TTACCGACGG CGATGATACG CCGTGGTCAA AGGCGTTCCT CGCCTCGGCC TACGCCTCCC 7800  
GCGGGTTGAA AATGCGCTAC ACCTCCGGCA CCGGATCCGA AGCGCTGATG GGCTATTTCG 7860  
AGAGCAAGTC GATGCTCTAC CTCGAATCGC GGTGCATCTT CATTACTAAA GGCGCCGGGG 7920  
TTCAGGGACT GCAAAACGGC GCGGTGAGCT GTATCGGCAT GACCGGCGCT GTGCCGTCCG 7980  
GCATTCGGGC GGTGCTGGCG GAAAACCTGA TCGCCTCTAT GCTCGACCTC GAAGTGGCGT 8040  
CCGCCAACGA CCAGACTTTC TCCCACTCGG ATATTCGCCG CACCGCGCGC ACCCTGATGC 8100  
AGATGCTGCC GGGCACCGAC TTTATTTTCT CCGGTACAG CGCGGTGCCG AACTACGACA 8160  
ACATGTTTCG CGGCTCGAAC TTCGATGCGG AAGATTTTGA TGATTACAAC ATCCTGCAGC 8220

-- 85 --

GTGACCTGAT GGTGACGGC GGCCTGCGTC CGGTGACCGA GCGGAAACC ATTGCCATTC	8280
GCCAGAAAGC GCGCGGGCG ATCCAGGCGG TTTTCCGCGA GCTGGGGCTG CCGCCAATCG	8340
CCGACGAGGA GGTGGAGGCC GCCACCTACG CGCACGGCAG CAACGAGATG CCGCCGCGTA	8400
ACGTGGTGGG GGATCTGAGT GCGGTGGAAG AGATGATGAA GCGCAACATC ACCGGCCTCG	8460
ATATTGTCGG CGCGCTGAGC CGCAGCGGCT TTGAGGATAT CGCCAGCAAT ATTCTCAATA	8520
TGCTGCGCCA GCGGGTCACC GCGGATTACC TGCAGACCTC GGCCATTCTC GATCGGCAGT	8580
TCGAGGTGGT GAGTGCGGTC AACGACATCA ATGACTATCA GGGGCCGGGC ACCGGCTATC	8640
GCATCTCTGC CGAACGCTGG GCGGAGATCA AAAATATTCC GGGCGTGGTT CAGCCCAGCA	8700
CCATTGAATA AGGCGGTATT CCTGTGCAAC AGACAACCCA AATTCAGCCC TCTTTTACCC	8760
TGAAAACCCG CGAGGGCGGG GTAGCTTCTG CCGATGAACG CGCCGATGAA GTGGTGATCG	8820
GCGTCGGCCC TGCCTTCGAT AACACCAGC ATCACACTCT GATCGATATG CCCCATGGCG	8880
CGATCCTCAA AGAGCTGATT GCCGGGGTGG AAGAAGAGGG GCTTCACGCC CGGGTGGTGC	8940
GCATTCTGCG CACGTCCGAC GTCTCCTTTA TGGCCTGGGA TCGGGCCAAC CTGAGCGGCT	9000
CGGGGATCGG CATCGGTATC CAGTCGAAGG GGACCACGGT CATCCATCAG CGCGATCTGC	9060
TGCCGCTCAG CAACCTGGAG CTGTTCTCCC AGGCGCCGCT GCTGACGCTG GAGACCTACC	9120
GGCAGATTGG CAAAAACGCT GCGCGCTATG CGCGCAAAGA GTCACCTTCG CCGGTGCCGG	9180
TGGTGAACGA TCAGATGGTG CGGCCGAAAT TTATGGCCAA AGCCGCGCTA TTTCATATCA	9240
AAGAGACCAA ACATGTGGTG CAGGACGCCG AGCCCGTCAC CCTGCACATC GACTTAGTAA	9300
GGGAGTGACC ATGAGCGAGA AAACCATGCG CGTGCAAGAT TATCCGTTAG CCACCCGCTG	9360
CCCGGAGCAT ATCCTGACGC CTACCGGCAA ACCATTGACC GATATTACCC TCGAGAAGGT	9420
GCTCTCTGGC GAGGTGGGCC CGCAGGATGT GCGGATCTCC CGCCAGACCC TTGAGTACCA	9480
GGCGCAGATT GCCGAGCAGA TGCAGCGCCA TGCGGTGGCG CGCAATTTCC GCCGCGCGGC	9540
GGAGCTTATC GCCATTCTTG ACGAGCGCAT TCTGGCTATC TATAACGCGC TCGCCCCGTT	9600
CCGCTCCTCG CAGGCGGAGC TGCTGGCGAT CGCCGACGAG CTGGAGCACA CCTGGCATGC	9660
GACAGTGAAT GCCGCCTTTG TCCGGGAGTC GCGGGAAGTG TATCAGCAGC GGCATAAGCT	9720
GCGTAAAGGA AGCTAAGCGG AGGTCAGCAT GCCGTTAATA GCCGGGATTG ATATCGGCAA	9780
CGCCACCACC GAGGTGGCGC TGGCGTCCGA CTACCCGCGC GCGAGGGCGT TTGTTGCCAG	9840
CGGGATCGTC GCGACGACGG GCATGAAAGG GACGCGGGAC AATATCGCCG GGACCCTCGC	9900

-- 86 --

CGCGCTGGAG CAGGCCCTGG CGAAAACACC GTGGTCGATG AGCGATGTCT CTCGCATCTA 9960  
TCTTAACGAA GCCGCGCCGG TGATTGGCGA TGTGGCGATG GAGACCATCA CCGAGACCAT 10020  
TATCACCGAA TCGACCATGA TCGGTCATAA CCCGCAGACG CCGGGCGGGG TGGGCGTTGG 10080  
CGTGGGGACG ACTATCGCCC TCGGGCGGCT GCGGACGCTG CCGGCGGCGC AGTATGCCGA 10140  
GGGGTGGATC GTACTGATTG ACGACGCCGT CGATTTCCTT GACGCCGTGT GGTGGCTCAA 10200  
TGAGGCGCTC GACCGGGGGA TCAACGTGGT GCGGCGGATC CTCAAAAAGG ACGACGGCGT 10260  
GCTGGTGAAC AACCGCCTGC GTAAAACCTT GCCGGTGGTG GATGAAGTGA CGCTGCTGGA 10320  
GCAGGTCCCC GAGGGGGTAA TGGCGGCGGT GGAAGTGGCC GCGCCGGGCC AGGTGGTGCG 10380  
GATCCTGTCTG AATCCCTACG GGATCGCCAC CTTCTTCGGG CTAAGCCCGG AAGAGACCCA 10440  
GGCCATCGTC CCCATCGCCC GCGCCCTGAT TGGCAACCGT TCCGCGGTGG TGCTCAAGAC 10500  
CCCGCAGGGG GATGTGCAGT CGCGGGTGAT CCCGGCGGGC AACCTCTACA TTAGCGGCGA 10560  
AAAGCGCCGC GGAGAGGCCG ATGTCGCCGA GGGCGCGGAA GCCATCATGC AGGCGATGAG 10620  
CGCCTGCGCT CCGGTACGCG ACATCCGCGG CGAACCGGGC ACCCACGCCG GCGGCATGCT 10680  
TGAGCGGGTG CGCAAGGTAA TGGCGTCCCT GACCGGCCAT GAGATGAGCG CGATATACAT 10740  
CCAGGATCTG CTGGCGGTGG ATACGTTTAT TCCGCGCAAG GTGCAGGGCG GGATGGCCGG 10800  
CGAGTGCGCC ATGGAGAATG CCGTCGGGAT GCGGCGGATG GTGAAAGCGG ATCGTCTGCA 10860  
AATGCAGGTT ATCGCCCGCG AACTGAGCGC CCGACTGCAG ACCGAGGTGG TGGTGGGCGG 10920  
CGTGGAGGCC AACATGGCCA TCGCCGGGGC GTTAACCACT CCCGGCTGTG CGGCGCCGCT 10980  
GGCGATCCTC GACCTCGGCG CCGGCTCGAC GGATGCGGCG ATCGTCAACG CGGAGGGGCA 11040  
GATAACGGCG GTCCATCTCG CCGGGGCGGG GAATATGGTC AGCCTGTTGA TTAAAACCGA 11100  
GCTGGGCCTC GAGGATCTTT CGCTGGCGGA AGCGATAAAA AAATACCCGC TGGCCAAAGT 11160  
GGAAAGCCTG TTCAGTATTC GTCACGAGAA TGGCGCGGTG GAGTTCTTTC GGAAGCCCT 11220  
CAGCCCGGCG GTGTTGCGCA AAGTGGTGTA CATCAAGGAG GCGGAACTGG TGCCGATCGA 11280  
TAACGCCAGC CCGCTGGAAA AAATTCGTCT CGTGCGCCGG CAGGCGAAAAG AGAAAGTGTT 11340  
TGTCACCAAC TGCCTGCGCG CGCTGCGCCA GGTCTACCC GCGGTTCCA TTCGCGATAT 11400  
CGCCTTTGTG GTGCTGGTGG GCGGCTCATC GCTGGACTTT GAGATCCCGC AGCTTATCAC 11460  
GGAAGCCTTG TCGCACTATG GCGTGGTCGC CGGGCGGGC AATATTCGGG GAACAGAAGG 11520  
GCCGCGCAAT GCGGTCGCCA CCGGGCTGCT ACTGGCCGGT CAGGCGAATT AAACGGGCGC 11580

-- 87 --

TCGCGCCAGC CTCTCTCTTT AACGTGCTAT TTCAGGATGC CGATAATGAA CCAGACTTCT 11640  
 ACCTTAACCG GGCAGTGCCT GGCCGAGTTT CTTGGCACCG GATTGCTCAT TTTCTTCGGC 11700  
 GCGGGTCTGCG TCGCTGCGCT GCGGGTCGCC GGGGCCAGCT TTGGTCAGTG GGAGATCAGT 11760  
 ATTATCTGGG GCCTTGGCGT CGCCATGGCC ATCTACCTGA CGGCCGGTGT CTCCGGCGCG 11820  
 CACCTAAATC CGGCGGTGAC CATTGCCCTG TGGCTGTTCG CCTGTTTTGA ACGCCGCAAG 11880  
 GTGCTGCCGT TTATTGTTGC CCAGACGGCC GGGGCCTTCT GCGCCGCCGC GCTGGTGTAT 11940  
 GGGCTCTATC GCCAGCTGTT TCTCGATCTT GAACAGAGTC AGCATATCGT GCGCGGCACT 12000  
 GCCGCCAGTC TTAACCTGGC CGGGGTCTTT TCCACGTACC CGCATCCACA TATCACTTTT 12060  
 ATACAAGCGT TTGCCGTGGA GACCACCATC ACGGCAATCC TGATGGCGAT GATCATGGCC 12120  
 CTGACCGACG ACGGCAACGG AATTC 12145

## (2) INFORMATION FOR SEQ ID NO:20:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 94 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

AGCTTAGGAG TCTAGAATAT TGAGCTCGAA TTCCCGGGCA TGCGGTACCG GATCCAGAAA 60  
 AAAGCCCGCA CCTGACAGTG CGGGCTTTTT TTTT 94

## (2) INFORMATION FOR SEQ ID NO:21:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 37 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (genomic)

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGAATTCAGA TCTCAGCAAT GAGCGAGAAA ACCATGC 37

## (2) INFORMATION FOR SEQ ID NO:22:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid

-- 88 --

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GCTCTAGATT AGCTTCCTTT ACGCAGC

27

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 33 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GGCCAAGCTT AAGGAGGTTA ATTAAATGAA AAG

33

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 26 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GCTCTAGATT ATTCAATGGT GTCGGG

26

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 42 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GCGCCGTCTA GAATTATGAG CTATCGTATG TTTGATTATC TG

42

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 base pairs
  - (B) TYPE: nucleic acid

-- 89 --

- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TCTGATACGG GATCCTCAGA ATGCCTGGCG GAAAT

36

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 51 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCGCGGATCC AGGAGTCTAG AATTATGGGA TTGACTACTA AACCTCTATC T

51

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 36 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GATACGCCCG GGTACCATT TCAACAGATC GTCCTT

36

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

TCGACGAATT CAGGAGGA

18

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 18 base pairs
  - (B) TYPE: nucleic acid

-- 90 --

(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CTAGTCCTCC TGAATTCG

18

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CTAGTAAGGA GGACAATTC

19

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

CATGGAATTG TCCTCCTTA

19

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 271 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: unknown  
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:  
(A) ORGANISM: GPPI

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Met Lys Arg Phe Asn Val Leu Lys Tyr Ile Arg Thr Thr Lys Ala Asn  
1                      5                      10                      15

Ile Gln Thr Ile Ala Met Pro Leu Thr Thr Lys Pro Leu Ser Leu Lys



-- 91 --

20	25	30
Ile Asn Ala Ala Leu Phe Asp Val Asp Gly Thr Ile Ile Ile Ser Gln		
35	40	45
Pro Ala Ile Ala Ala Phe Trp Arg Asp Phe Gly Lys Asp Lys Pro Tyr		
50	55	60
Phe Asp Ala Glu His Val Ile His Ile Ser His Gly Trp Arg Thr Tyr		
65	70	75
Asp Ala Ile Ala Lys Phe Ala Pro Asp Phe Ala Asp Glu Glu Tyr Val		
85	90	95
Asn Lys Leu Glu Gly Glu Ile Pro Glu Lys Tyr Gly Glu His Ser Ile		
100	105	110
Glu Val Pro Gly Ala Val Lys Leu Cys Asn Ala Leu Asn Ala Leu Pro		
115	120	125
Lys Glu Lys Trp Ala Val Ala Thr Ser Gly Thr Arg Asp Met Ala Lys		
130	135	140
Lys Trp Phe Asp Ile Leu Lys Ile Lys Arg Pro Glu Tyr Phe Ile Thr		
145	150	155
Ala Asn Asp Val Lys Gln Gly Lys Pro His Pro Glu Pro Tyr Leu Lys		
165	170	175
Gly Arg Asn Gly Leu Gly Phe Pro Ile Asn Glu Gln Asp Pro Ser Lys		
180	185	190
Ser Lys Val Val Val Phe Glu Asp Ala Pro Ala Gly Ile Ala Ala Gly		
195	200	205
Lys Ala Ala Gly Cys Lys Ile Val Gly Ile Ala Thr Thr Phe Asp Leu		
210	215	220
Asp Phe Leu Lys Glu Lys Gly Cys Asp Ile Ile Val Lys Asn His Glu		
225	230	235
Ser Ile Arg Val Gly Glu Tyr Asn Ala Glu Thr Asp Glu Val Glu Leu		
245	250	255
Ile Phe Asp Asp Tyr Leu Tyr Ala Lys Asp Asp Leu Leu Lys Trp		
260	265	270

## (2) INFORMATION FOR SEQ ID NO:34:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 555 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

## (ii) MOLECULE TYPE: protein

-- 92 --

(vi) ORIGINAL SOURCE:

(A) ORGANISM: DHAB1

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```

Met Lys Arg Ser Lys Arg Phe Ala Val Leu Ala Gln Arg Pro Val Asn
1           5           10           15

Gln Asp Gly Leu Ile Gly Glu Trp Pro Glu Glu Gly Leu Ile Ala Met
          20           25           30

Asp Ser Pro Phe Asp Pro Val Ser Ser Val Lys Val Asp Asn Gly Leu
          35           40           45

Ile Val Glu Leu Asp Gly Lys Arg Arg Asp Gln Phe Asp Met Ile Asp
          50           55           60

Arg Phe Ile Ala Asp Tyr Ala Ile Asn Val Glu Arg Thr Glu Gln Ala
65           70           75           80

Met Arg Leu Glu Ala Val Glu Ile Ala Arg Met Leu Val Asp Ile His
          85           90           95

Val Ser Arg Glu Glu Ile Ile Ala Ile Thr Thr Ala Ile Thr Pro Ala
          100          105          110

Lys Ala Val Glu Val Met Ala Gln Met Asn Val Val Glu Met Met Met
          115          120          125

Ala Leu Gln Lys Met Arg Ala Arg Arg Thr Pro Ser Asn Gln Cys His
          130          135          140

Val Thr Asn Leu Lys Asp Asn Pro Val Gln Ile Ala Ala Asp Ala Ala
145          150          155          160

Glu Ala Gly Ile Arg Gly Phe Ser Glu Gln Glu Thr Thr Val Gly Ile
          165          170          175

Ala Arg Tyr Ala Pro Phe Asn Ala Leu Ala Leu Leu Val Gly Ser Gln
          180          185          190

Cys Gly Arg Pro Gly Val Leu Thr Gln Cys Ser Val Glu Glu Ala Thr
          195          200          205

Glu Leu Glu Leu Gly Met Arg Gly Leu Thr Ser Tyr Ala Glu Thr Val
          210          215          220

Ser Val Tyr Gly Thr Glu Ala Val Phe Thr Asp Gly Asp Asp Thr Pro
225          230          235          240

Trp Ser Lys Ala Phe Leu Ala Ser Ala Tyr Ala Ser Arg Gly Leu Lys
          245          250          255

Met Arg Tyr Thr Ser Gly Thr Gly Ser Glu Ala Leu Met Gly Tyr Ser
          260          265          270

```

-- 93 --

Glu Ser Lys Ser Met Leu Tyr Leu Glu Ser Arg Cys Ile Phe Ile Thr  
 275 280 285  
 Lys Gly Ala Gly Val Gln Gly Leu Gln Asn Gly Ala Val Ser Cys Ile  
 290 295 300  
 Gly Met Thr Gly Ala Val Pro Ser Gly Ile Arg Ala Val Leu Ala Glu  
 305 310 315 320  
 Asn Leu Ile Ala Ser Met Leu Asp Leu Glu Val Ala Ser Ala Asn Asp  
 325 330 335  
 Gln Thr Phe Ser His Ser Asp Ile Arg Arg Thr Ala Arg Thr Leu Met  
 340 345 350  
 Gln Met Leu Pro Gly Thr Asp Phe Ile Phe Ser Gly Tyr Ser Ala Val  
 355 360 365  
 Pro Asn Tyr Asp Asn Met Phe Ala Gly Ser Asn Phe Asp Ala Glu Asp  
 370 375 380  
 Phe Asp Asp Tyr Asn Ile Leu Gln Arg Asp Leu Met Val Asp Gly Gly  
 385 390 395 400  
 Leu Arg Pro Val Thr Glu Ala Glu Thr Ile Ala Ile Arg Gln Lys Ala  
 405 410 415  
 Ala Arg Ala Ile Gln Ala Val Phe Arg Glu Leu Gly Leu Pro Pro Ile  
 420 425 430  
 Ala Asp Glu Glu Val Glu Ala Ala Thr Tyr Ala His Gly Ser Asn Glu  
 435 440 445  
 Met Pro Pro Arg Asn Val Val Glu Asp Leu Ser Ala Val Glu Glu Met  
 450 455 460  
 Met Lys Arg Asn Ile Thr Gly Leu Asp Ile Val Gly Ala Leu Ser Arg  
 465 470 475 480  
 Ser Gly Phe Glu Asp Ile Ala Ser Asn Ile Leu Asn Met Leu Arg Gln  
 485 490 495  
 Arg Val Thr Gly Asp Tyr Leu Gln Thr Ser Ala Ile Leu Asp Arg Gln  
 500 505 510  
 Phe Glu Val Val Ser Ala Val Asn Asp Ile Asn Asp Tyr Gln Gly Pro  
 515 520 525  
 Gly Thr Gly Tyr Arg Ile Ser Ala Glu Arg Trp Ala Glu Ile Lys Asn  
 530 535 540  
 Ile Pro Gly Val Val Gln Pro Asp Thr Ile Glu  
 545 550 555

-- 94 --

## (2) INFORMATION FOR SEQ ID NO:35:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 194 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

## (vi) ORIGINAL SOURCE:

(A) ORGANISM: DHAB2

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```

Met Gln Gln Thr Thr Gln Ile Gln Pro Ser Phe Thr Leu Lys Thr Arg
1           5           10           15

Glu Gly Gly Val Ala Ser Ala Asp Glu Arg Ala Asp Glu Val Val Ile
20           25           30

Gly Val Gly Pro Ala Phe Asp Lys His Gln His His Thr Leu Ile Asp
35           40           45

Met Pro His Gly Ala Ile Leu Lys Glu Leu Ile Ala Gly Val Glu Glu
50           55           60

Glu Gly Leu His Ala Arg Val Val Arg Ile Leu Arg Thr Ser Asp Val
65           70           75           80

Ser Phe Met Ala Trp Asp Ala Ala Asn Leu Ser Gly Ser Gly Ile Gly
85           90           95

Ile Gly Ile Gln Ser Lys Gly Thr Thr Val Ile His Gln Arg Asp Leu
100          105          110

Leu Pro Leu Ser Asn Leu Glu Leu Phe Ser Gln Ala Pro Leu Leu Thr
115          120          125

Leu Glu Thr Tyr Arg Gln Ile Gly Lys Asn Ala Ala Arg Tyr Ala Arg
130          135          140

Lys Glu Ser Pro Ser Pro Val Pro Val Val Asn Asp Gln Met Val Arg
145          150          155          160

Pro Lys Phe Met Ala Lys Ala Ala Leu Phe His Ile Lys Glu Thr Lys
165          170          175

His Val Val Gln Asp Ala Glu Pro Val Thr Leu His Ile Asp Leu Val
180          185          190

Arg Glu

```

-- 95 --

## (2) INFORMATION FOR SEQ ID NO:36:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 140 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: DHAB3

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```

Met Ser Glu Lys Thr Met Arg Val Gln Asp Tyr Pro Leu Ala Thr Arg
1           5           10           15
Cys Pro Glu His Ile Leu Thr Pro Thr Gly Lys Pro Leu Thr Asp Ile
          20           25           30
Thr Leu Glu Lys Val Leu Ser Gly Glu Val Gly Pro Gln Asp Val Arg
          35           40           45
Ile Ser Arg Gln Thr Leu Glu Tyr Gln Ala Gln Ile Ala Glu Gln Met
          50           55           60
Gln His Ala Val Ala Arg Asn Phe Arg Arg Ala Ala Glu Leu Ile Ala
65           70           75           80
Ile Pro Asp Glu Arg Ile Leu Ala Ile Tyr Asn Ala Leu Arg Pro Phe
          85           90           95
Arg Ser Ser Gln Ala Glu Leu Leu Ala Ile Ala Asp Glu Leu Glu His
          100          105          110
Thr Trp His Ala Thr Val Asn Ala Ala Phe Val Arg Glu Ser Ala Glu
          115          120          125
Val Tyr Gln Gln Arg His Lys Leu Arg Lys Gly Ser
          130          135          140

```

## (2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 387 amino acids  
 (B) TYPE: amino acid  
 (C) STRANDEDNESS: unknown  
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(vi) ORIGINAL SOURCE:  
 (A) ORGANISM: DHAT

-- 96 --

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

```

Met Ser Tyr Arg Met Phe Asp Tyr Leu Val Pro Asn Val Asn Phe Phe
1           5           10           15
Gly Pro Asn Ala Ile Ser Val Val Gly Glu Arg Cys Gln Leu Leu Gly
          20           25           30
Gly Lys Lys Ala Leu Leu Val Thr Asp Lys Gly Leu Arg Ala Ile Lys
          35           40           45
Asp Gly Ala Val Asp Lys Thr Leu His Tyr Leu Arg Glu Ala Gly Ile
          50           55           60
Glu Val Ala Ile Phe Asp Gly Val Glu Pro Asn Pro Lys Asp Thr Asn
65           70           75           80
Val Arg Asp Gly Leu Ala Val Phe Arg Arg Glu Gln Cys Asp Ile Ile
          85           90           95
Val Thr Val Gly Gly Gly Ser Pro His Asp Cys Gly Lys Gly Ile Gly
          100          105          110
Ile Ala Ala Thr His Glu Gly Asp Leu Tyr Gln Tyr Ala Gly Ile Glu
          115          120          125
Thr Leu Thr Asn Pro Leu Pro Pro Ile Val Ala Val Asn Thr Thr Ala
          130          135          140
Gly Thr Ala Ser Glu Val Thr Arg His Cys Val Leu Thr Asn Thr Glu
145          150          155          160
Thr Lys Val Lys Phe Val Ile Val Ser Trp Arg Lys Leu Pro Ser Val
          165          170          175
Ser Ile Asn Asp Pro Leu Leu Met Ile Gly Lys Pro Ala Ala Leu Thr
          180          185          190
Ala Ala Thr Gly Met Asp Ala Leu Thr His Ala Val Glu Ala Tyr Ile
          195          200          205
Ser Lys Asp Ala Asn Pro Val Thr Asp Ala Ala Ala Met Gln Ala Ile
          210          215          220
Arg Leu Ile Ala Arg Asn Leu Arg Gln Ala Val Ala Leu Gly Ser Asn
225          230          235          240
Leu Gln Ala Arg Glu Asn Met Ala Tyr Ala Ser Leu Leu Ala Gly Met
          245          250          255
Ala Phe Asn Asn Ala Asn Leu Gly Tyr Val His Ala Met Ala His Gln
          260          265          270
Leu Gly Gly Leu Tyr Asp Met Pro His Gly Val Ala Asn Ala Val Leu
          275          280          285

```

-- 97 --

Leu Pro His Val Ala Arg Tyr Asn Leu Ile Ala Asn Pro Glu Lys Phe  
 290 295 300

Ala Asp Ile Ala Glu Leu Met Gly Glu Asn Ile Thr Gly Leu Ser Thr  
 305 310 315 320

Leu Asp Ala Ala Glu Lys Ala Ile Ala Ala Ile Thr Arg Leu Ser Met  
 325 330 335

Asp Ile Gly Ile Pro Gln His Leu Arg Asp Leu Gly Val Lys Glu Ala  
 340 345 350

Asp Phe Pro Tyr Met Ala Glu Met Ala Leu Lys Asp Gly Asn Ala Phe  
 355 360 365

Ser Asn Pro Arg Lys Gly Asn Glu Gln Glu Ile Ala Ala Ile Phe Arg  
 370 375 380

Gln Ala Phe  
 385

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GCGAATTCAT GAGCTATCGT ATGTTTG

27

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

GCGAATTCAG AATGCCTGGC GGAAAATC

28

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

-- 98 --

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GGGAATTCAT GAGCGAGAAA ACCATGCG

28

(2) INFORMATION FOR SEQ ID NO:41:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GCGAATTCTT AGCTTCCTTT ACGCAGC

27

(2) INFORMATION FOR SEQ ID NO:42:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GCGAATTCAT GCAACAGACA ACCCAAATTC

30

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

GCGAATTCAC TCCCTTACTA AGTCG

25

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear



-- 99 --

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

GGGAATTCAT GAAAAGATCA AAACGATTTG

30

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

GCGAATTCTT ATTCAATGGT GTCGGGCTG

29

(2) INFORMATION FOR SEQ ID NO:46

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 34 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

TTGATAATAT AACCATGGCT GCTGCTGCTG ATAG

34

(2) INFORMATION FOR SEQ ID NO:47

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

GTATGATATG TTATCTTGA TCCAATAAAT CTAATCTTC

39

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single

-- 100 --

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CATGACTAGT AAGGAGGACA ATTC

24

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CATGGAATTG TCCTCCTTAC TAGT

24

## WHAT IS CLAIMED IS:

1. An improved method for the production of 1,3-propanediol from an organism capable of producing 1,3-propanediol, said organism comprising at least one gene encoding a dehydratase activity, the method comprising the steps of:
  - (a) introducing a gene encoding protein X into the organism to create a transformed organism; and
  - (b) culturing the transformed organism in the presence of at least one carbon source capable of being converted to 1,3 propanediol in said transformed host organism and under conditions suitable for the production of 1,3 propanediol wherein the carbon source is selected from the group consisting of monosaccharides, oligosaccharides, polysaccharides, and a one carbon substrate.
2. The method of Claim 1 further comprising the step of introducing at least one gene encoding a protein selected from the group consisting of protein 1, protein 2 and protein 3 into the organism.
3. The method of Claim 1 further comprising the step of recovering the 1,3 propanediol.
4. The method of Claim 1 wherein the gene encoding protein X is isolated from a glycerol dehydratase gene cluster.
5. The method of Claim 1 wherein the gene encoding protein X is isolated from a diol dehydratase gene cluster.
6. The method of Claim 4 wherein the glycerol dehydratase gene cluster is from an organism selected from the genera consisting of *Klebsiella* and *Citrobactor*.
7. The method of Claim 5 wherein the diol dehydratase gene cluster is from an organism selected from the genera consisting of *Klebsiella*, *Clostridium* and *Salmonella*.
8. The method of Claim 1 wherein the gene encoding a dehydratase activity is heterologous to the organism.
9. The method of Claim 1 wherein the gene encoding a dehydratase activity is homologous to the organism.

-- 102 --

10. The method of Claim 1 wherein the organism is selected from the group of genera consisting of *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*, *Bacillus*, *Streptomyces* and *Pseudomonas*.
11. The method of Claim 10 wherein the organism is selected from the group consisting of *E.coli* and *Klebsiella spp.*
12. The method of Claim 1 wherein the gene encoding protein X is stably maintained in the host genome.
13. The method of Claim 2 wherein at least one gene encoding a protein selected from protein 1, protein 2 and protein 3 is stably maintained in the host genome.
14. The method of Claim 1 wherein the carbon source is glucose.
15. The method of Claim 1 wherein the gene encoding protein X has the sequence as shown in SEQ ID NO: 59.
16. The method of Claim 2 wherein protein 1 has the sequence as shown in SEQ ID NO: 60 or SEQ ID NO: 61.
17. The method of Claim 2 wherein protein 2 has the sequence as shown in SEQ ID NO: 62 or SEQ ID NO: 63.
18. The method of Claim 2 wherein protein 3 has the sequence as shown in SEQ ID NO: 64 or SEQ ID NO: 65.
19. A recombinant microorganism capable of producing 1,3-propanediol from a carbon source said recombinant microorganism comprising a) at least one gene encoding a dehydratase activity; b) at least one gene encoding a glycerol-3-phosphatase; and c) at least one gene encoding protein X.
20. The recombinant microorganism of Claim 19 further comprising d) at least one gene encoding a protein selected from the group consisting of protein 1, protein 2 and protein 3.

21. The recombinant microorganism of Claim 19 selected from the group consisting of *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*, *Bacillus*, *Streptomyces* and *Pseudomonas*.
22. The recombinant microorganism of Claim 19 wherein the gene encoding protein X is isolated from a glycerol dehydratase gene cluster.
23. The recombinant microorganism of Claim 19 wherein the gene encoding protein X is isolated from a diol dehydratase gene cluster.
24. The recombinant microorganism of Claim 22 wherein the glycerol dehydratase gene cluster is from an organism selected from the genera consisting of *Klebsiella* and *Citrobacter*.
25. The recombinant microorganism of Claim 23 wherein the diol dehydratase gene cluster is from an organism selected from the genera consisting of *Klebsiella*, *Clostridium* and *Salmonella*.
26. The recombinant microorganism of Claim 19 wherein said dehydratase activity is heterologous to said microorganism.
27. The recombinant microorganism of Claim 19 wherein said dehydratase activity is homologous to said microorganism.
28. The recombinant microorganism of Claim 19 wherein the gene encoding protein X has the sequence as shown in SEQ ID NO: 59.
29. The recombinant microorganism of Claim 20 wherein protein 1 has the sequence as shown in SEQ ID NO: 60 or SEQ ID NO: 61.
30. The method of Claim 20 wherein protein 2 has the sequence as shown in SEQ ID NO: 62 or SEQ ID NO: 63.

31. The method of Claim 20 wherein protein 3 has the sequence as shown in SEQ ID: 64 or SEQ ID NO: 65.
32. A method for extending the half-life of dehydratase activity in a microorganism capable of producing 1,3-propanediol and containing at least one gene encoding a dehydratase activity, comprising the step of introducing a gene encoding protein X into said microorganism and culturing under conditions suitable for production of 1,3-propanediol.
33. The method of Claim 32 wherein the gene encoding the dehydratase activity is heterologous to said microorganism.
34. The method of Claim 32 wherein the gene encoding the dehydratase activity is homologous to said microorganism.
35. The microorganism of Claim 32 wherein the gene encoding protein X is isolated from a glycerol dehydratase gene cluster.
36. The microorganism of Claim 32 wherein the gene encoding protein X is isolated from a diol dehydratase gene cluster.
37. The microorganism of Claim 35 wherein the glycerol dehydratase gene cluster is from an organism selected from the genera consisting of *Klebsiella* and *Citrobacter*.
38. The microorganism of Claim 34 wherein the diol dehydratase gene cluster is from an organism selected from the genera consisting of *Klebsiella*, *Clostridium* and *Salmonella*.
39. The method of Claim 32 wherein the microorganism is selected from the group consisting of *Citrobacter*, *Enterobacter*, *Clostridium*, *Klebsiella*, *Aerobacter*, *Lactobacillus*, *Aspergillus*, *Saccharomyces*, *Schizosaccharomyces*, *Zygosaccharomyces*, *Pichia*, *Kluyveromyces*, *Candida*, *Hansenula*, *Debaryomyces*, *Mucor*, *Torulopsis*, *Methylobacter*, *Escherichia*, *Salmonella*, *Bacillus*, *Streptomyces* and *Pseudomonas*.
40. The method of Claim 32 further comprising the step of introducing a gene encoding at least one of protein 1, protein 2 and protein 3 into said microorganism.

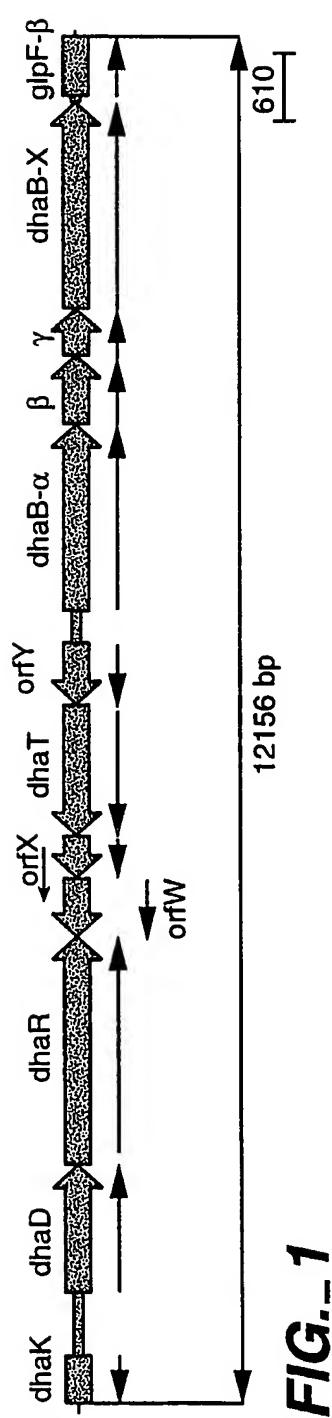


FIG.\_1

	M	S	L	S	S	P	G	V	H	L	F	Y	D	S	R	G	Q	G	A	G	A	L	D	E	L	C	W	G	L	E	E	Q	G	V	P	C	Q	A	I	T	Majority	
	10																																									
1	M	S	L	S	S	P	G	V	H	L	F	Y	H	S	R	W	Q	G	T	R	V	L	D	E	L	C	W	G	L	E	E	Q	G	V	P	C	R	A	I	C	cfu_orfX.aa	
1	M	S	L	S	P	P	G	V	R	L	F	Y	D	P	R	G	H	A	G	A	I	N	E	L	C	W	G	L	E	E	Q	G	V	P	C	Q	T	I	T	kpn_orfX.aa		
	50																																									
	Y	D	G	G	G	D	A	A	L	G	A	L	A	K	S	S	T	L	R	V	G	L	G	L	S	A	S	G	D	I	A	L	T	H	A	Q	L	P	Majority			
	60																																									
41	C	D	D	H	D	C	A	L	A	L	G	K	L	A	A	K	S	S	T	L	R	V	G	L	G	L	N	A	T	G	D	I	A	L	T	H	A	Q	L	P	cfu_orfX.aa	
41	Y	D	G	G	G	D	A	A	L	G	A	L	A	A	R	S	S	P	L	R	V	G	I	G	L	S	A	S	G	E	I	A	L	T	H	A	Q	L	P	kpn_orfX.aa		
	90																																									
	A	D	A	A	L	A	T	G	H	V	T	A	G	T	A	Q	L	R	T	L	G	A	N	A	G	Q	L	V	K	V	L	P	L	S	E	R	I	K	Majority			
	100																																									
81	E	D	R	A	L	V	C	G	H	T	R	A	G	T	A	Q	I	R	T	L	G	A	N	A	G	Q	L	V	K	V	L	P	F	S	E	-	I	K	cfu_orfX.aa			
81	A	D	A	P	L	A	T	G	H	V	T	D	S	D	D	Q	L	R	T	L	G	A	N	A	G	Q	L	V	K	V	L	P	L	S	E	R	N	.	kpn_orfX.aa			

FIG.\_4

2 / 27

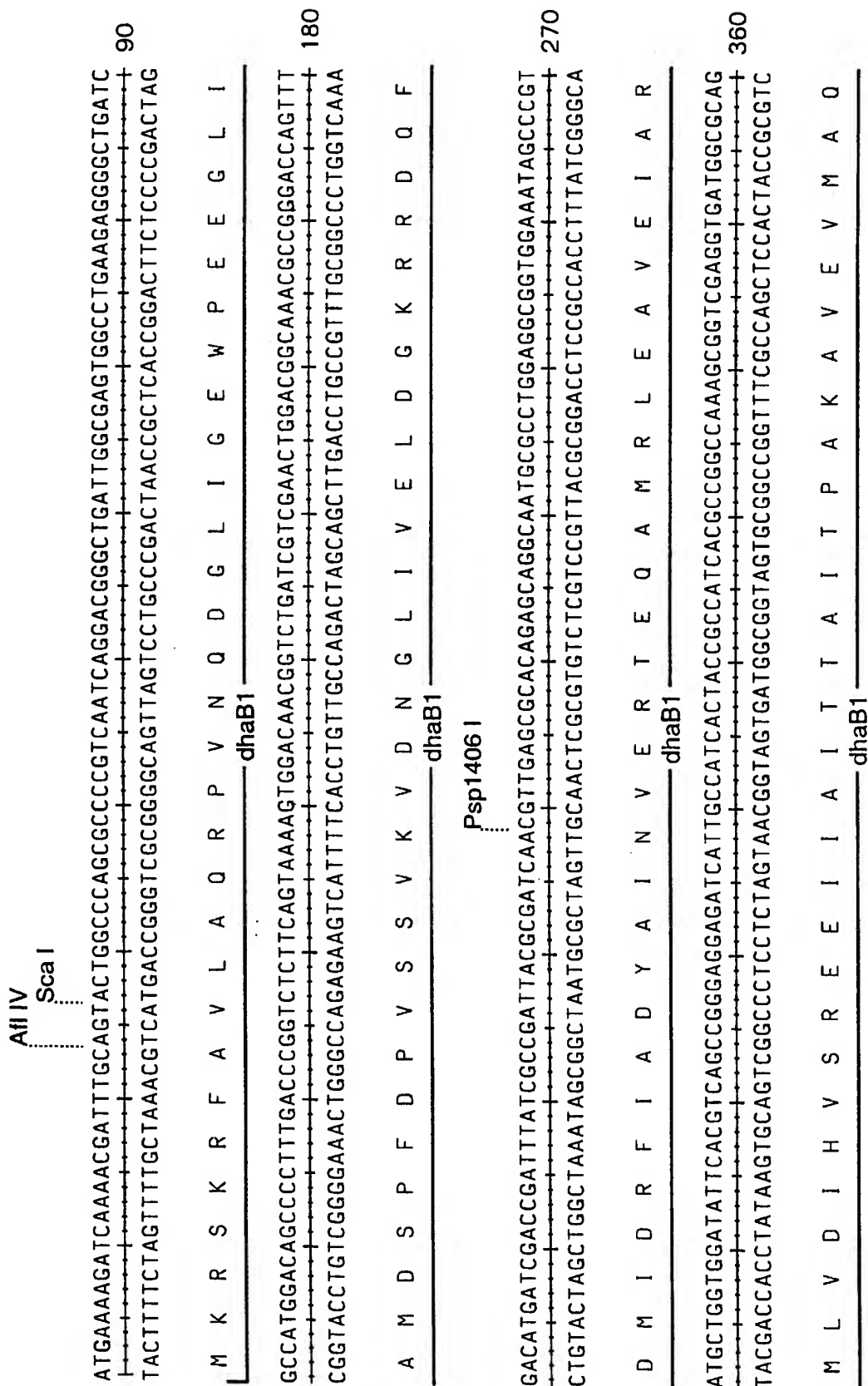
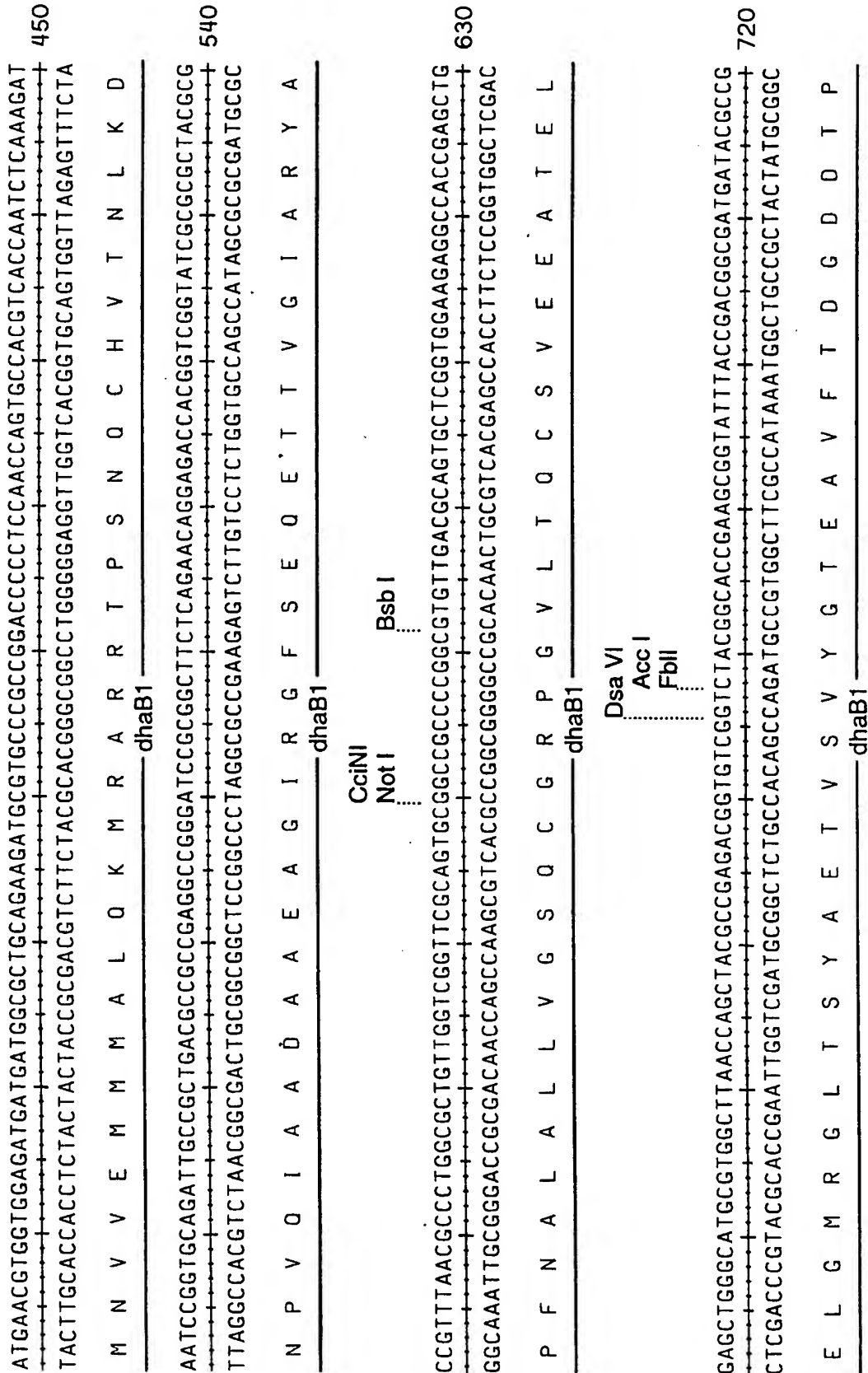


FIG.\_2A-1



3/27

**FIG.\_2A-2**

4/27

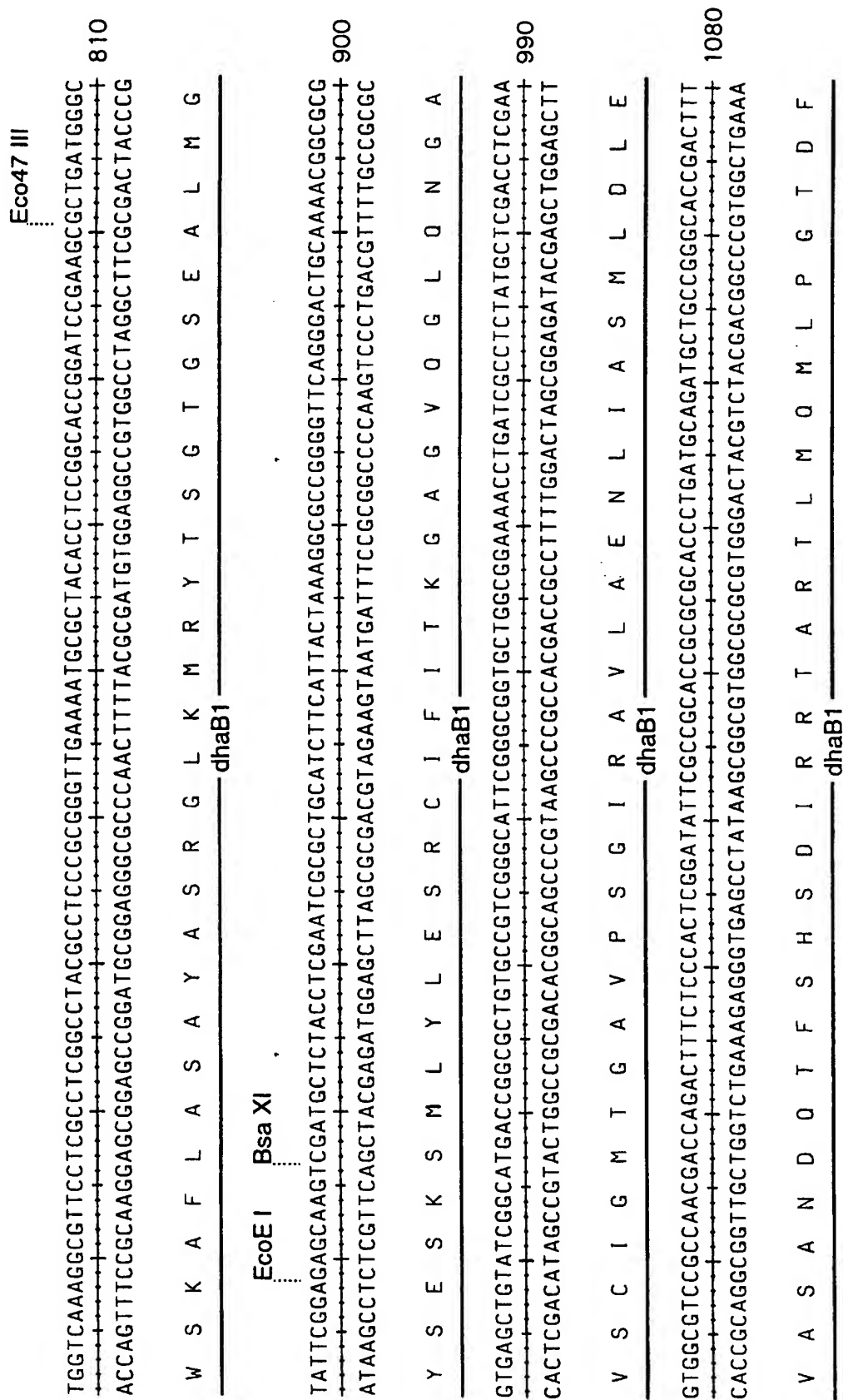


FIG. 2B-1

5/27

Pfl11081

ATTTTCTCGGCTACAGCGGGTGCCGAACIACGACAACATGTTCCGCCGCTCGAACTTCGATCGGAAGATTTTGAIGATTACAACATC  
 TAAAGAGGCCGATGTCGCCACCGCTTGATGCTGTGTACAGCGGCCGAGCTTGAAGCTACGCCCTTCTAAAACTACTAATGTTGTAG  
 1170

I F S G Y S A V P N Y D N M F A G S N F D A E D F D D Y N I  
 dhaB1

CTGCAGCGTGACCTGATGGTTGACGGCGGCTGCCGTCCGGTGACCGAGGGCGGAACCATTTGCCATTTCGCCAGAAAGCGGGCGGGCGATC  
 GACGTGCACTGGACTACCAACTGCCCGCGGACGAGGCCACTGGCTCCGCCCTTTGGTAACGGTAAGCGGTCTTTCGCCGCGGCCGCTAG  
 1260

L Q R D L M V D G G L R P V T E A E T I A I R Q K A A R A I  
 dhaB1

CAGGCGGTTTTCCGCGAGCTGGGGCTGCCGCCCAATCGCCGACGAGGAGGTGGAGGGCCGCCACCTACGGCGACGGCAGCAACGAGATGCCG  
 GTCCGCCAAAAGGCGCTCGACCCCGACGCGGTTAGCGGCTGCTCCTCCACCTCCGGCGGTGGATGCCGCTGCCGTCTGCTCTACGGC  
 1350

Q A V F R E L G L P P I A D E E V E A A T Y A H G S N E M P  
 dhaB1

CCGCGTAACGTGGTGGAGGATCTGAGTGCGGTGGAGAGAGATGATGAAGCGCAACATCACCGGCCCTCGATATTGTCGGCGCGCTGAGCCGC  
 GGCGCATTCACCACTCCTAGACTCAGCCACCTTCTCTACTACTTCCGCTGTAGTGCCGCGGAGCTATAACAGCCGCGGCGACTCGGCG  
 1440

P R N V V E D L S A V E E M M K R N I T G L D I V G A L S R  
 dhaB1

**FIG.-2B-2**

6/27

AGCGGCTTTGAGGATATCGCCAGCAATATTCTCAATATGCTGCGCCAGCGGGTCACCGGCGATTACCTGCAGACCTCGGCCATTCTCGAT  
 TCGCCGAAACTCCTATAGCGGTCGTTATAAGAGTTATACGACGCGGTGCGCCAGTGGCCGCTAATGGACGCTCTGGAGCCGGTAAGAGCTA  
 S G F E D I A S N I L N M L R O R V T G D Y L O T S A I L D  
 dhaB1 1530

FIG.\_2B-3

CGGCAGTTCGAGGTGGTGAGTGGGTCAACGACATCAATGACTATCAGGGGCGGCGCACCGGCTATCGCACTCTGCGGAACGCTGGGCG  
 GCCGTCAAGCTCCACCACTCAGCGCAGTTGCTGTAGTTACTGATAGTCCCGGCGCGGTGGCCGATAGCGTAGAGACGGCTTGGACCCGC  
 R O F E V V S A V N D I N D Y Q G P G T G Y R I S A E R W A  
 dhaB1 1620

Drd II

GAGATCAAAAATATTCCGGGGTGGTTCAGCCCGACACCATTTGAATAAGGCGGTATTCCTGTGCAACAGACAACCCAAATTCAGCCCTCT  
 CTCTAGTTTATAAGGCCCGCACCAAGTCGGGCTGTGGTAACCTATTCCGCCATAAGGACACGTTGCTGTTGGGTTTAAGTCGGGAGA  
 E I K N I P G V V Q P D T I E  
 dhaB1 1710

H<sub>[dh]</sub>

TTTACCCTGAAACCGCGAGGGGGGTAGCTTCTGCCGATGAACGCGCGCATGAAGTGGTGATCGGCGTCGGCCCTGCCCTTCGATAAA  
 AAATGGGACTTTTGGGCGCTCCCGCCCATCGAAGACGGCTACTTGGCGGCTACTTCACCACTAGCCGACGCCGGGACGGAAGCTATT  
 F T L K T R E G G V A S A D E R A D E V V I G V G P A F D K  
 dhaB2 1800

FIG.\_2C-1

7/27

Uba1220 I  
 Cfr9I  
 PspAI  
 Xma I  
 CfrJ4I  
 Sma I

CACCAGCATCACA CTGATCGATATGCCCCATGGCGGATCCTCAAGAGCTGATTGCCGGGGTGGAAGAGAGGGGCTTCACGCCCGG  
 GTGGTCGTAGTGTGAGACTAGCTATACGGGGTACCGCGCTAGGAGTTTCTCGACTAACGGCCCCACCTTCTTCTCCCCGAAGTGC GGCC  
 1890

H Q H H T L I D M P H G A I L K E L I A G V E E G L H A R  
 dhaB2

Ppu1253 I  
 Aat II

GTGGTGGCATTCTGGCACGTCGCGACGTCTCCTTTATGGCTGGGATGGGCCAACCTGAGCGGCTCGGGGATCGGCATCGGTATCCAG  
 CACCACGCGTAAGACGCGTGCAGGCTGCAGAGGAAATACCGGACCCTAGCCCGTTGGACTCGCCGAGCCCTAGCCGTAGCCATAGGTC  
 1980

V V R I L R T S D V S F M A W D A A N L S G S G I G I G I Q  
 dhaB2

TCGAAGGGGACCACGGTCATCCATCAGCGCGATCTGCTGCCGCTCAGCAACCTGGAGCTGTTCTCCAGGCGCGCTGCTGACGCTGGAG  
 AGCTTCCCCTGGTGCCAGTAGGTAGTCGCGCTAGACGACGCGGAGTCGTTGGACCTCGACAAGAGGGTCCGCGGCGACGACTGCGGACCTC  
 2070

S K G T T V I H O R D L L P L S N L E L F S O A P L L T L E  
 dhaB2

**FIG.\_2C-2**

8 / 27

ACCTACCGGCAGATTGGCAAAACGCTGCGCGCTATGCGCGCAAGAGTCACCTTCGCCGGTGCCGGTGGTGAACGATCAGATGGTGCGG 2160  
 TGGATGGCGGCTCTAACCGTTTTTGGAGCGCGGATACGCGCGTTTCTCAGTGAAGCGGCCACGGCCACCACTTGCTAGTCTACCAAGCC  
 T Y R Q I G K N A A R Y A R K E S P S P V V N D Q M V R  
 -----dhaB2

**FIG.\_2C-3**

CCGAAATTTATGGCCAAAGCCGCGCTATTTTCATATCAAGAGACCAAAACATGTGGTGCAGGACGCCGAGCCCGTCACCCCTGCACATCGAC 2250  
 GGCTTTAAATACCGGTTTCGGCGGATAAAGTATAGTTTCTCTGTTGTACACCACGTCCTCGGGCTCGGGCAGTGGGACGTGTAGCTG  
 P K F M A K A A L F H I K E T K H V V O D A E P V T L H I D  
 -----dhaB2

TTAGTAAGGGAGTGACCATGAGCGAGAAAACCATGCGGTGCAGGATTATCCGTTAGCCACCCGCTGCCCGGAGCATATCCTGACGCCTA 2340  
 AATCATCCCTCACTGGTACTCGCTCTTTGGTACGCGCACGTCCTAATAGGCAATCGGTGGGCGACGGGCCCTCGTATAGGACTGCGGAT  
 M S E K T M R V O D Y P L A T R C P E H I L T P  
 -----dhaB3

L V R E  
 -----dhaB2

**FIG.\_2D-1**

9 / 27

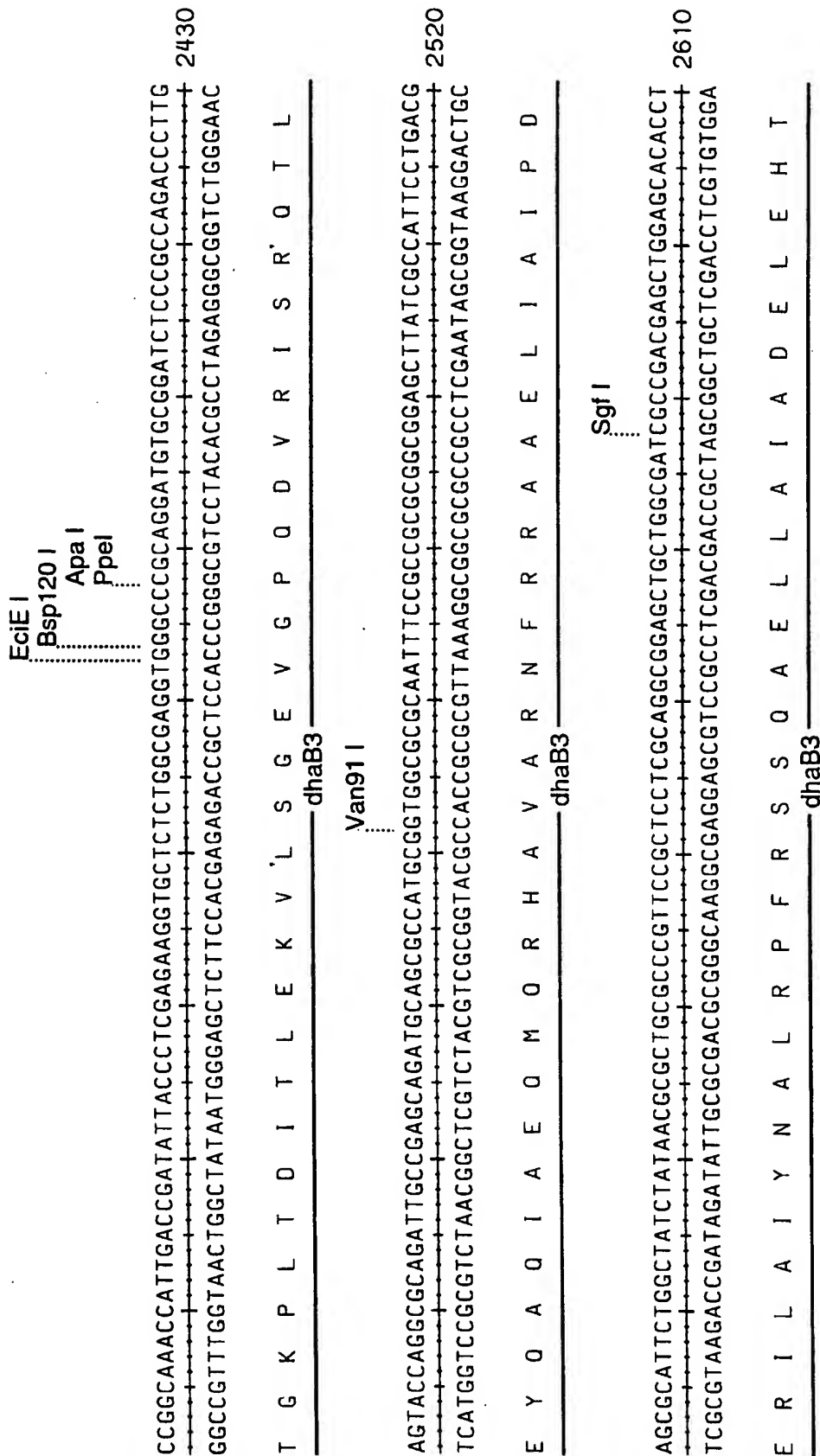


FIG.-2D-2

10 / 27

GGCATGCGACAGTGAATGCCGCTTIGTCGGGAGTCGGGGAAGTGATCAGCAGCGGCATAAGCTGCGTAAAGGAAGCTAAGCGGAGG 2700  
CCGTACGCTGTCACTTACGGCGGAACAGGCCCTCAGCCGCTTCACATAGTCGTGCGCGTATTCGACGCAATTCCTTCGATTCCGCTCC

W H A T V N A A F V R E S A E V Y Q Q R H K L R K G S  
-----dhaB3-----

Xcm I

TCAGCATGCCGTTAATAGCCGGGATTGATATCGGCAACGCCACCACCGAGGTGGCGTGGCGTCCGACTACCCGCAGCGAGGGCGTTTG 2790  
AGTCGTACGGCAATTATCGGCCCTAACTATAGCCGTTGCGGTGGTGGTCCACCGGACCGCAGGCTGATGGCGTCCGCTCCCGCAAAC

M P L I A G I D I G N A T T E V A L A S D Y P Q A R A F  
-----dhaB4-----

# FIG. 2D-3

SanD I

TTGCCAGCGGGATCGTCGCGACGACGGGCATGAAAGGACGCGGACAAATATCGCGGGACCCCTCGCCGCGCTGGAGCAGGCCCTGGCGGA 2880  
AACGGTCGCCCTAGCAGCGCTGCTGCCCGTACTTCCCTGCGCCCTGTTATAGCGGCCCTGGGAGCGGCGGACCTCGTCCGGGACCGCT

V A S G I V A T T G M K G T R D N I A G T L A A L E Q A L A  
-----dhaB4-----

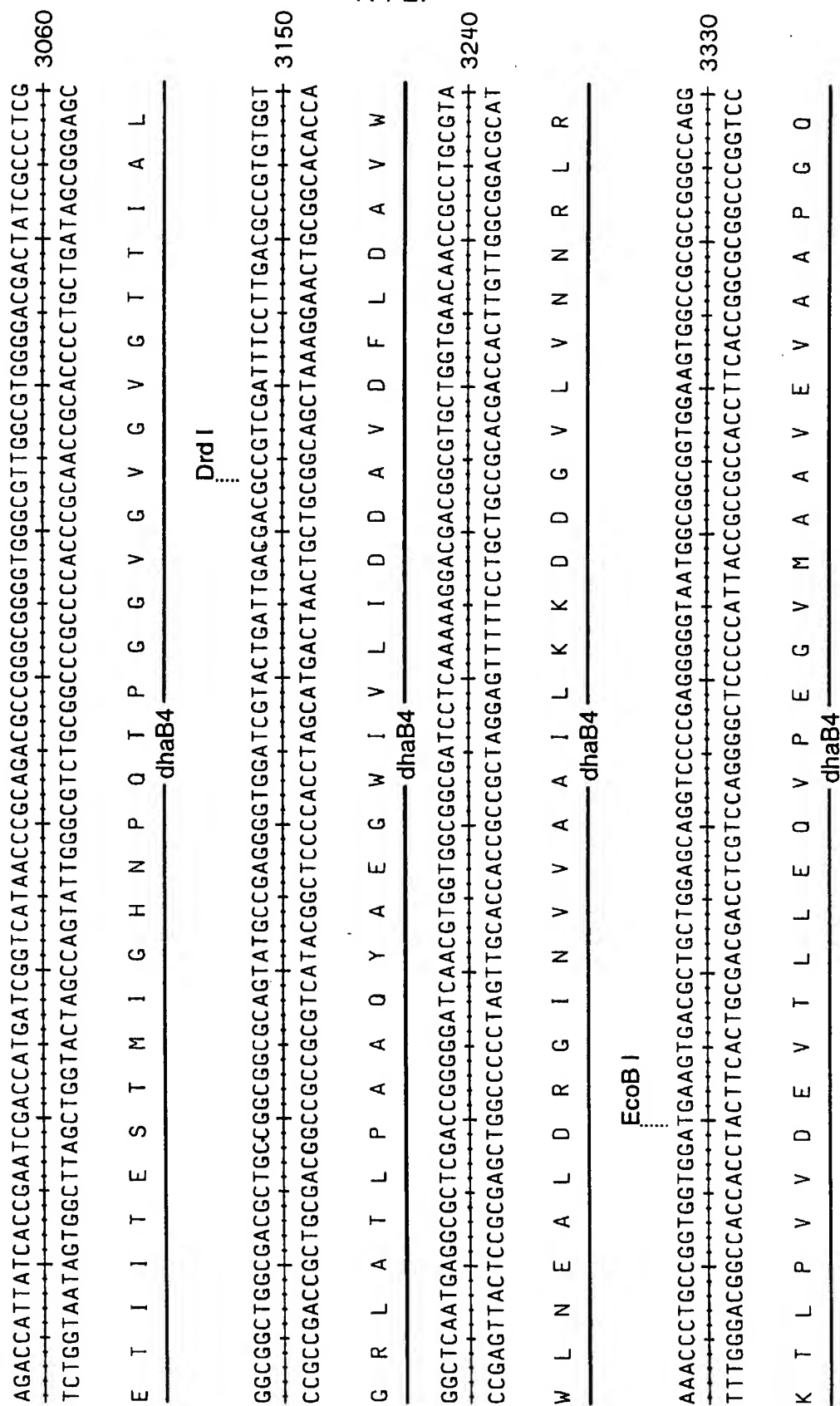
AAACACCGTGGTCGATGAGCGATGTCCTCGCATCTATCTTAACGAAGCCGCGCGGTGATTGGCGATGTGGCGATGGAGACCATCACCG 2970  
TTGTGCAACCACTACTCGCTACAGAGCGTAGATAGAATTGCTTCGGCGGGCCCACTAACCGCTACACCGCTACCTCTGGTAGTGGC

K T P W S M S D V S R I Y L N E A A P V I G D V A M E T I T  
-----dhaB4-----

# FIG. 2E-1



11 / 27

**FIG..2E-2**

TGGTGGGATCCIGTGAATCCCTACGGGATCGCCACCTTCTTCGGGCTAAGCCCGGAAGAGACCCAGGCCATCGTCCCCCATCGCCCGCG  
3420  
ACCACGCTAGGACAGCTTAGGGATGCCCTAGCGGTGAAGAAGCCGATTTCGGGCTTCTCTGGGTCCGGTAGCAGGGTAGCGGGCGC

V V R I L S N P Y G I A T F F G L S P E E T O A I V P I A R  
-----dhaB4-----

CCCTGATTGGCAACCGTTCGCGGTGGTGTCTCAAGACCCCGCAGGGGATGTGCAGTTCGCGGTGATCCCGGGGCAACCCTCTACATTA  
3510  
GGGACTAACCGTTGGCAAGCGCCACACGAGTTCTGGGGCTCCCCCTACACGTACAGTCCGCGCCCTAGGGCCGCGCTTGGAGATGTAAT

A L I G N R S A V V L K T P Q G D V Q S R V I P A G N L Y I  
-----dhaB4-----

GCGGCGAAAGCGCGGAGAGGCGGATGTGCGCGAGGGCGGAAGCCATCATGCAGGCGATGAGCGCTCCGGTACGCGACA  
3600  
CGCGGCTTTTCGCGGGCGCTCTCCGGCTACAGCGGCTCCCGGCTTCGGTAGTACGTCCGCTACTCGCGGACGCGAGGCCCATGCGCTGT

S G E K R R G E A D V A E G A E A I M O A M S A C A P V R D.  
-----dhaB4-----

**FIG.\_2E-3**

Taq II'

TCCGCGGCGAACCGGGCACCCACGCGGGCGGCATGCTTGAGCGGGTGCGCAAGGTAATGGCGTCCCTGACCGGCCATGAGATGAGCGCGA  
3690  
AGGCGCGCTTGGCCCGTGGGTGCGGCGCGCGTACGAACTCGCCCCACGCGTTCCATTACCGCAGGGAAGTGGCCGGTACTCTACTCGCGCT

I R G E P G T H A G G M L E R V R K V M A S L T G H E M S A  
-----dhaB4-----

**FIG.\_2F-1**

13 / 27

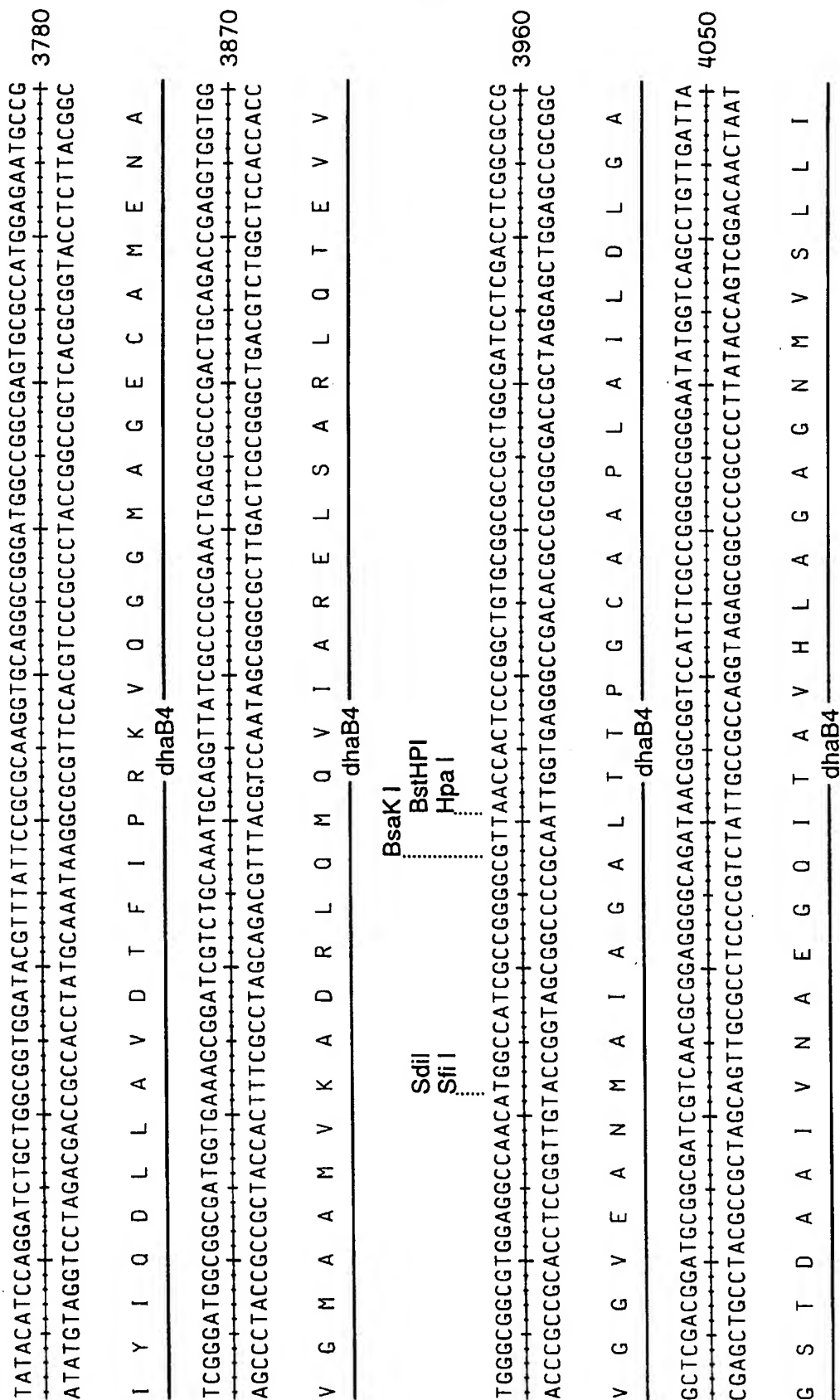


FIG. 2F-2

14/27

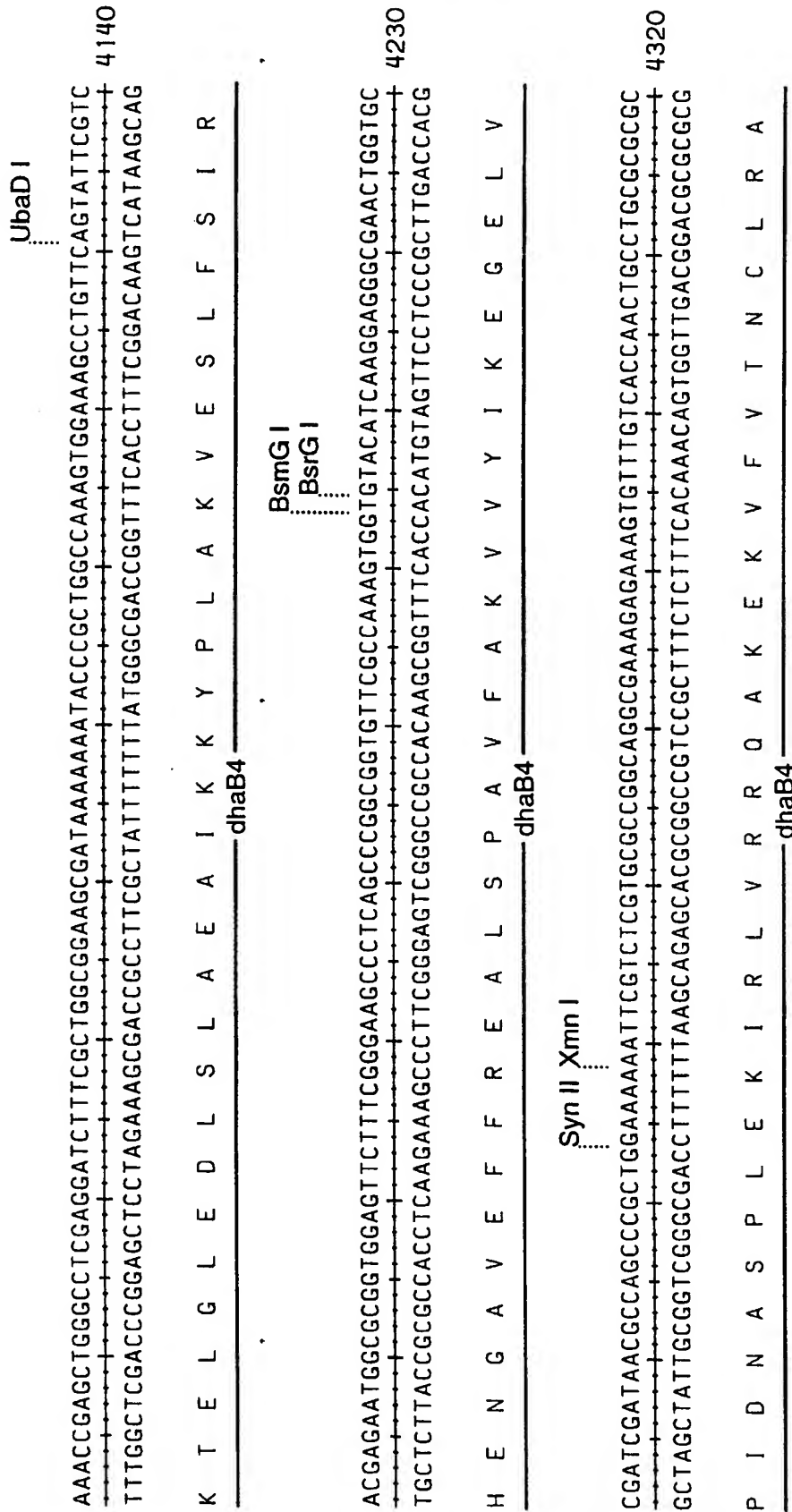
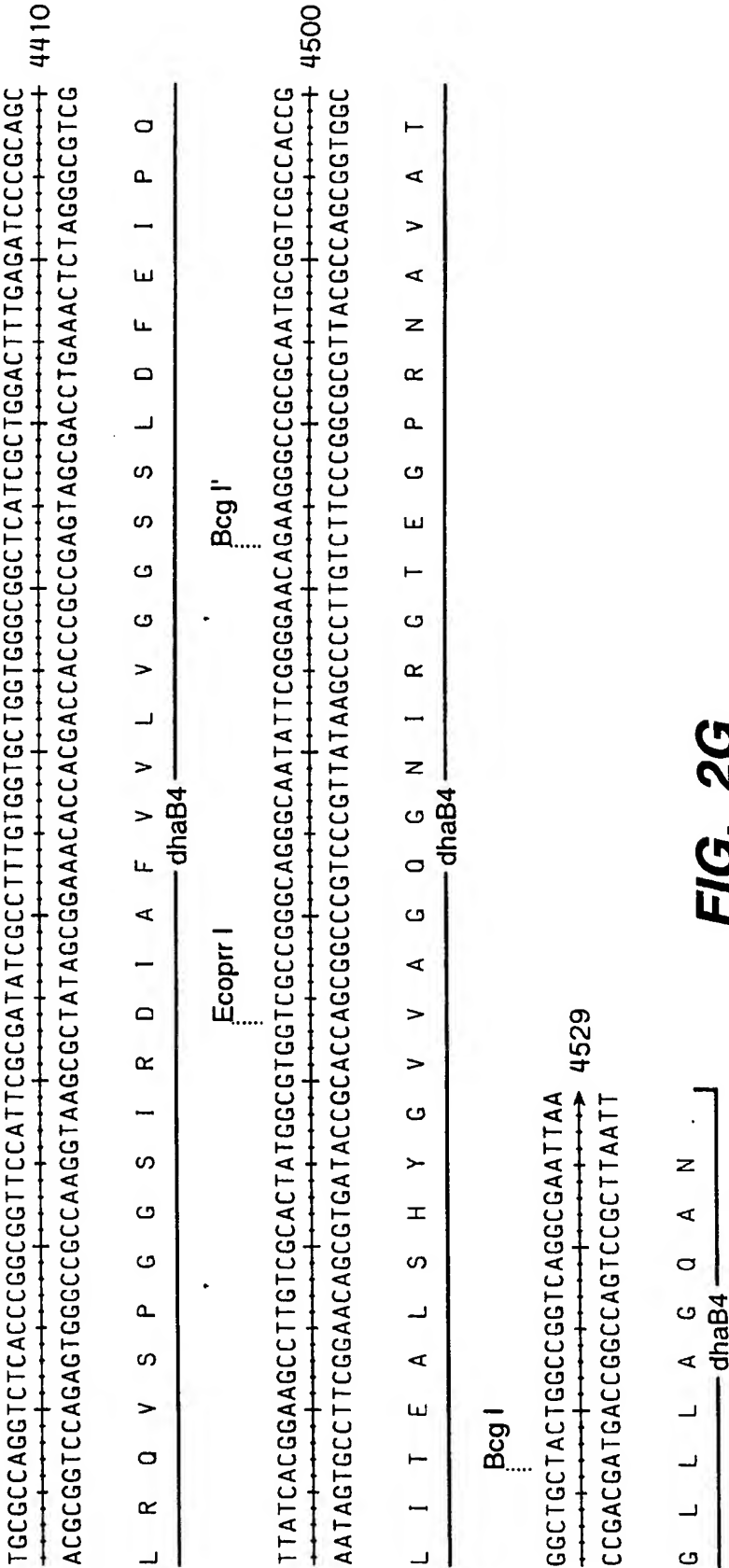


FIG. 2F-3



16/27

	M	M	N	K	S	Q	Q	V	A	T	I	T	L	A	A	Q	Q	M	A	A	V	E	A	K	A	L	E	I	N	V	A	V	F	S	V	D	Majority				
1	M	-	N	K	S	Q	Q	I	A	T	I	T	L	A	A	A	K	K	M	A	Q	A	V	E	A	K	A	L	E	I	N	V	P	V	F	S	V	D	cfu_orfy.aa		
1	M	M	N	K	S	Q	Q	V	Q	T	I	T	L	A	A	A	Q	Q	M	A	A	V	E	K	K	A	T	E	I	N	V	A	V	F	S	V	D	Kpn_orfy.aa			
	H	G	G	N	T	L	L	I	Q	R	M	D	D	A	F	V	S	S	C	D	I	S	L	N	K	A	Y	S	A	C	S	L	K	Q	G	T	H	E	I	T	Majority
40	H	G	G	N	T	L	L	M	Q	R	M	D	D	A	F	V	T	S	C	D	I	S	L	N	K	A	Y	T	A	C	C	L	R	Q	G	T	H	E	I	T	cfu_orfy.aa
41	R	G	G	N	T	L	L	I	Q	R	M	D	E	A	F	V	S	S	C	D	I	S	L	N	K	A	W	S	A	C	S	L	K	Q	G	T	H	E	I	T	Kpn_orfy.aa
	S	A	V	Q	P	G	A	S	L	Y	G	L	Q	L	T	N	Q	Q	R	I	V	I	F	G	G	G	L	P	V	I	L	N	G	Q	V	I	G	A	V	G	Majority

FIG.\_3

17/27

	MYRIYTRTGDN	GT	TALFGGSR	IDKDDIR	VEAYGT	VD	ELIS	Majority	
	10		20		30		40		
1	MYRIYTRTGDN	GT	TALFGGSR	IDKDDIR	VEAYGT	VD	ELIS	cfu_orfW.aa	
1	MYRIYTRTGDK	GT	TALYGGSR	IEKDHIR	VEAYGT	VD	ELIS	kpn_orfW.aa	
	QLGVCYASTR	DAGLR	ESLHAI	QQTLF	VLGAEL	ASDA	KGLT	Majority	
	50		60		70		80		
41	QLGVCYASTR	QAELRQ	ELHAMQ	KMLFV	LGAE	LASD	QKGLT	cfu_orfW.aa	
41	QLGVCYATTR	DAGLR	ESLHHI	QQTLF	VLGAEL	ASDA	RGLT	kpn_orfW.aa	
	RLSQTIGEE	DI	TAL	EQ	LIDR	NMAES	GPLKE	FVIPGKNLAS	Majority
	90		100		110		120		
81	RLKQRIGEE	DI	QAL	EQ	LIDR	NMAES	GPLKE	FVIPGKNLAS	cfu_orfW.aa
81	RLSQTIGEE	DI	TAL	ERL	LIDR	NMAES	GPLKE	FVIPGRNLAS	kpn_orfW.aa
	AQLHVARTL	SRRL	ERLL	IAMG	RALT	LRDA	AKRY	INRLSDA	Majority
	130		140		150		160		
121	AQLHVARTL	TRRL	ERLL	IAMG	RTRTL	LRDE	ARRY	INRLSDA	cfu_orfW.aa
121	AQLHVARTQ	SRRL	ERLL	TAMD	RAHPL	RDA	LKRY	SNRLSDA	kpn_orfW.aa
	LFSMARIEE	ET	TPDACA	-				Majority	
	170								
161	LFSMARIEE	ET	TPDVCA					cfu_orfW.aa	
161	LFSMARIEE	ET	RPDACA	.				kpn_orfW.aa	

FIG.-5

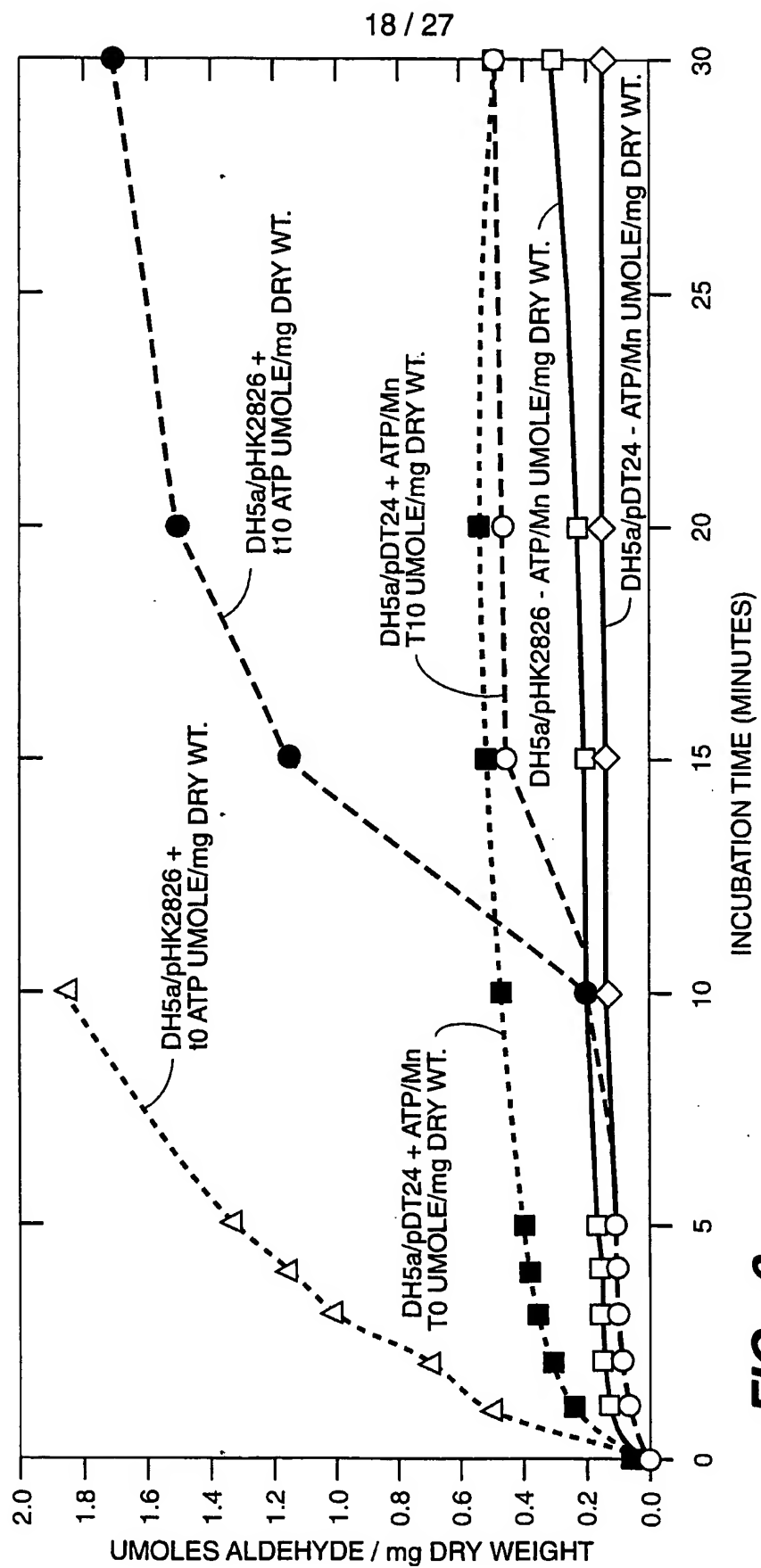


FIG. 6



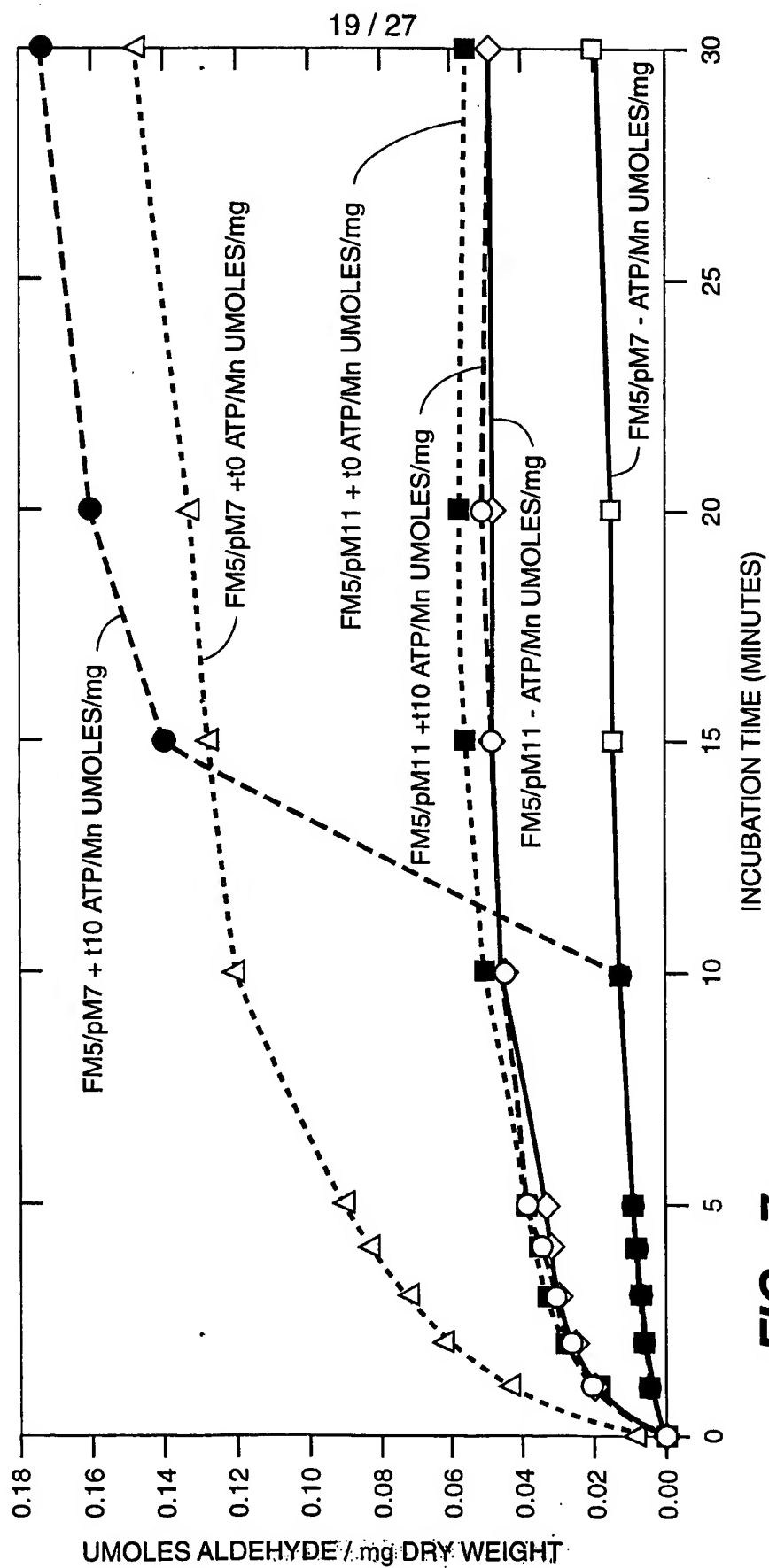


FIG. 7

20 / 27

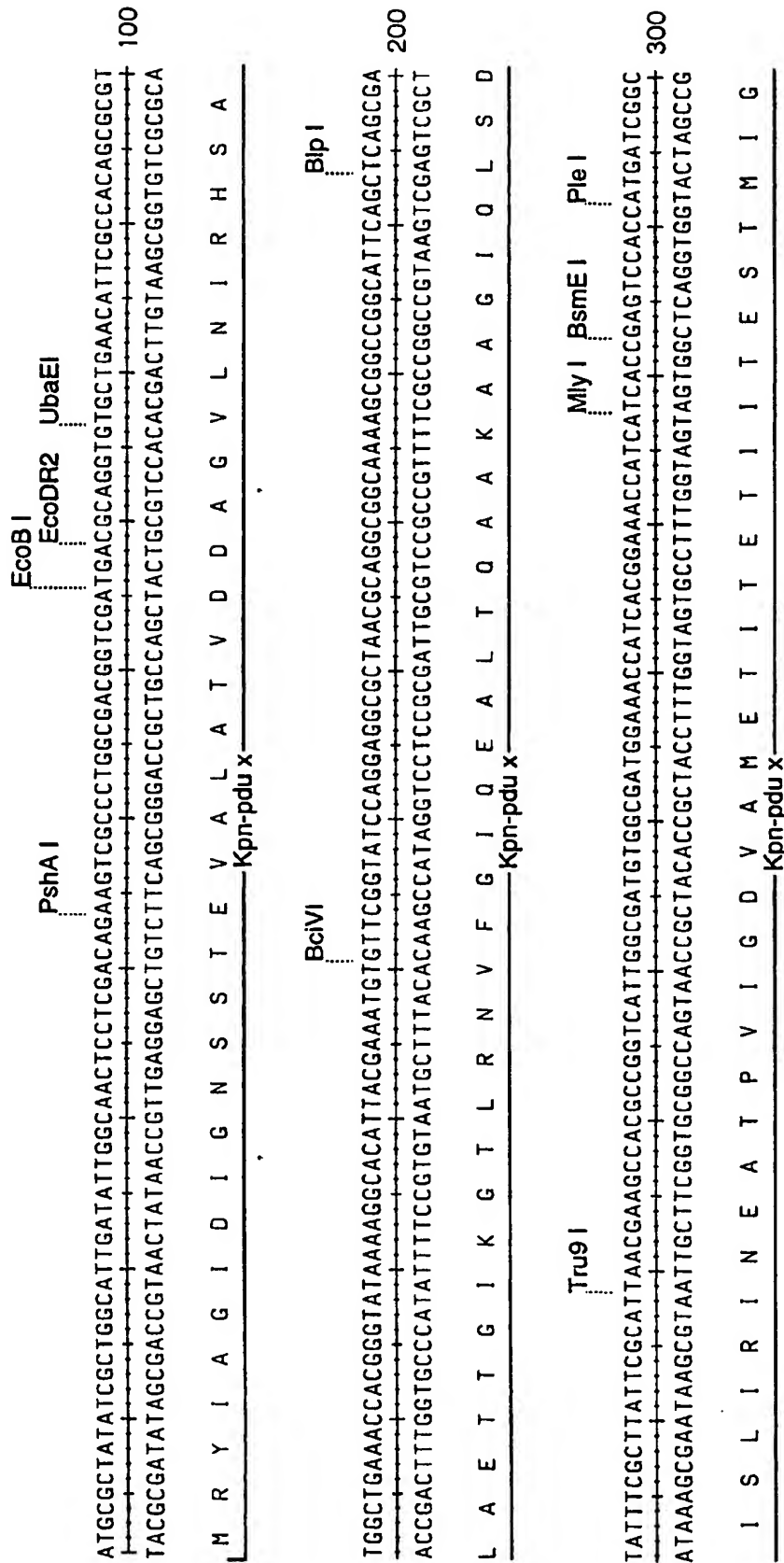


FIG.-8A-1

21 / 27

TGGTCTCCTCGGCCTTTGACCTTTGCGGATGTGCGCGCGATGGTCAATGCGGCAACGGCAGCGGGCTATCAGATAACCGGCATTATTTTGCAGCAGGATGA  
 ACCAGAGGAGCGCGAAACTGAAACGGCTACAGCGGCGCTACCAAGTTACGCCGTGCGCGGATAGTCTATTGGCCGTAATAAAACGTCGTCCTACT  
 V V S S A F D F A D V A A M V N A A T A A G Y Q I T G I I L Q Q D D  
 —————Kpn-pdu x—————

Age I  
PvuII

CjeP I

Tfi I

CjeP I'

CGGCGTGTGTTCAATAACCGGCTACAGCAACCGTACCGGTGATCGACGAAGTTCAGCATATCGACCGGATTCACATTGGCATGCTGGCGGCCGTCGAG

GCCGCACGACCAGTTATTGGCCGATGTCGTGGCGATGGCCCACTAGCTGCTCAAGTCGTATAGCTGGCCTAAGGTGAACCGTACGACCGCCGGCAGCTC

600

G V L V N N R L Q Q P L P V I D E V Q H I D R I P L G M L A A V E

Kpn-pdu x

**FIG. 8B-1**

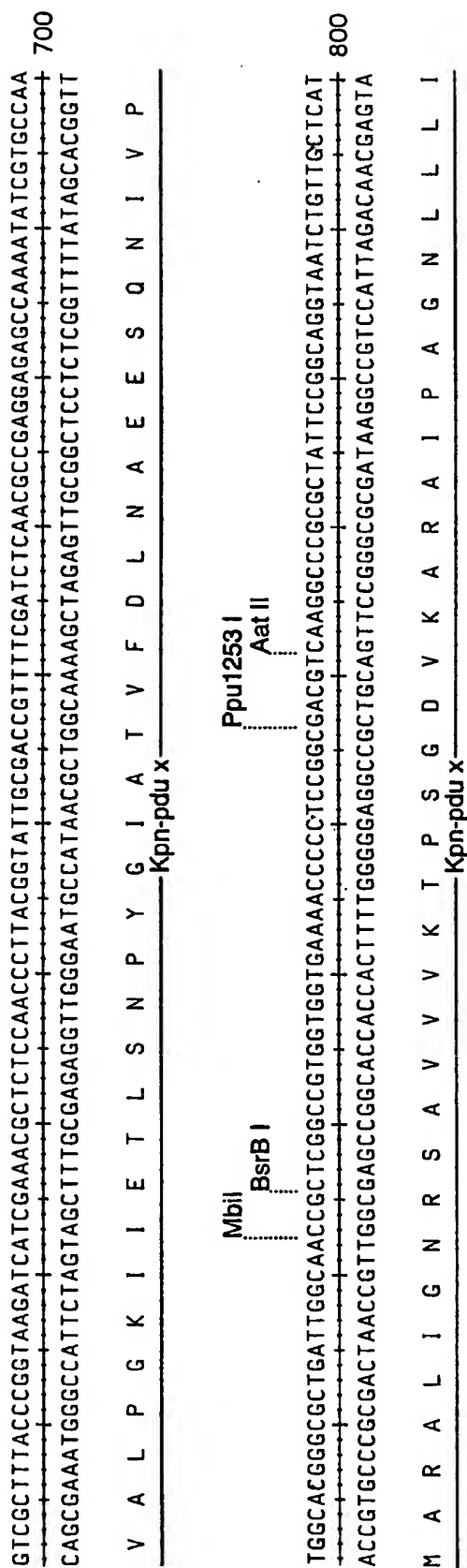


FIG.-8B-2

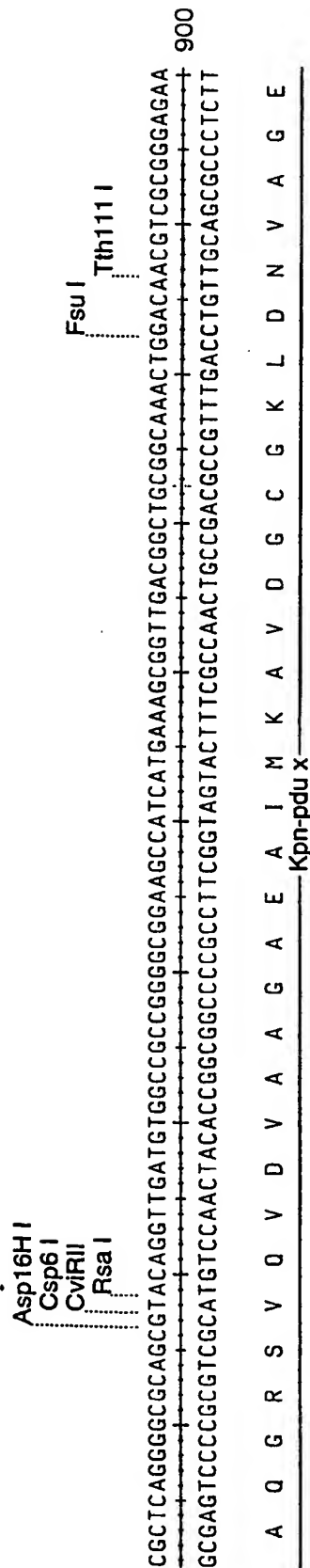


FIG.-8C-1

23 / 27

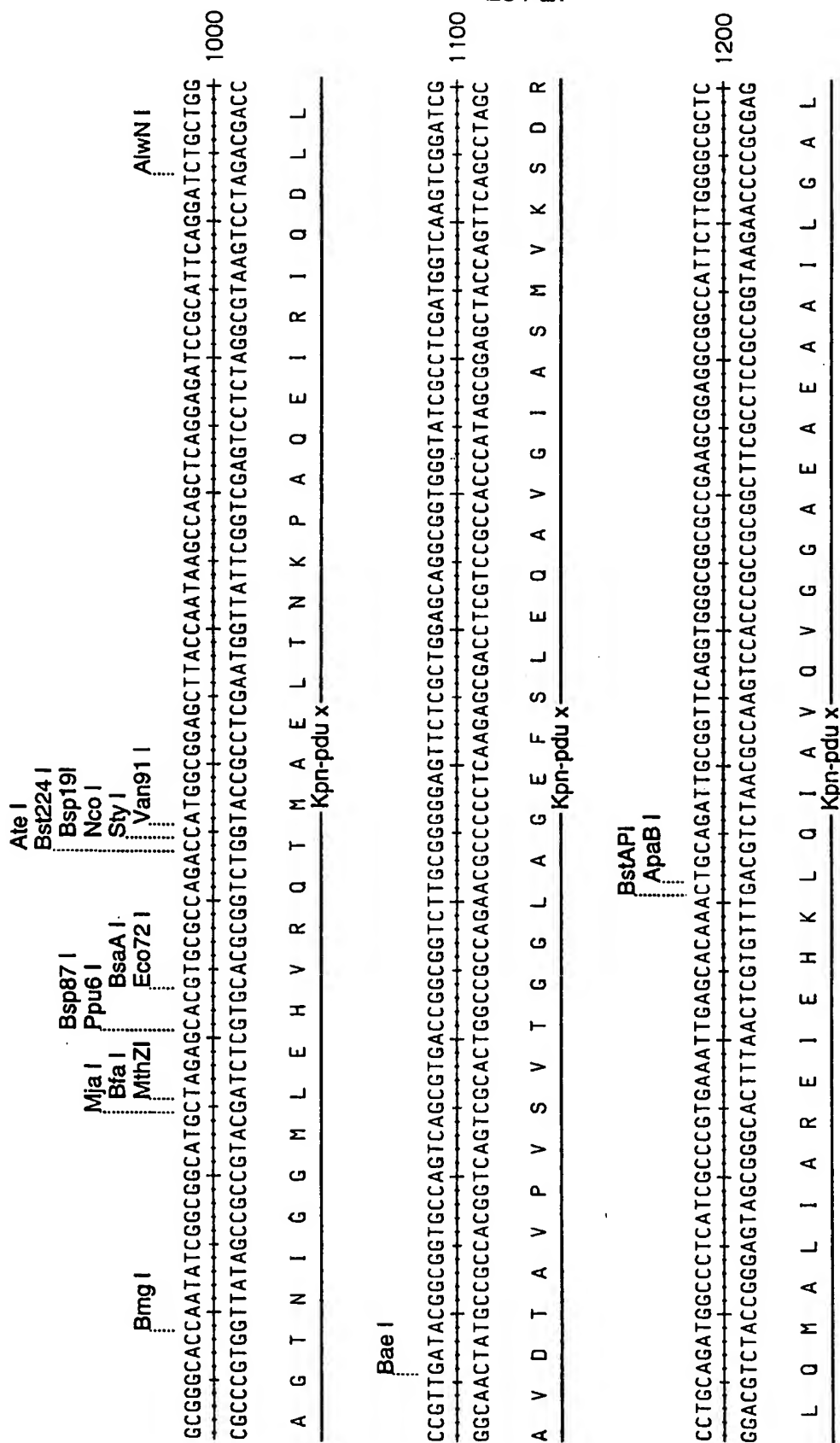


FIG. 8C-2

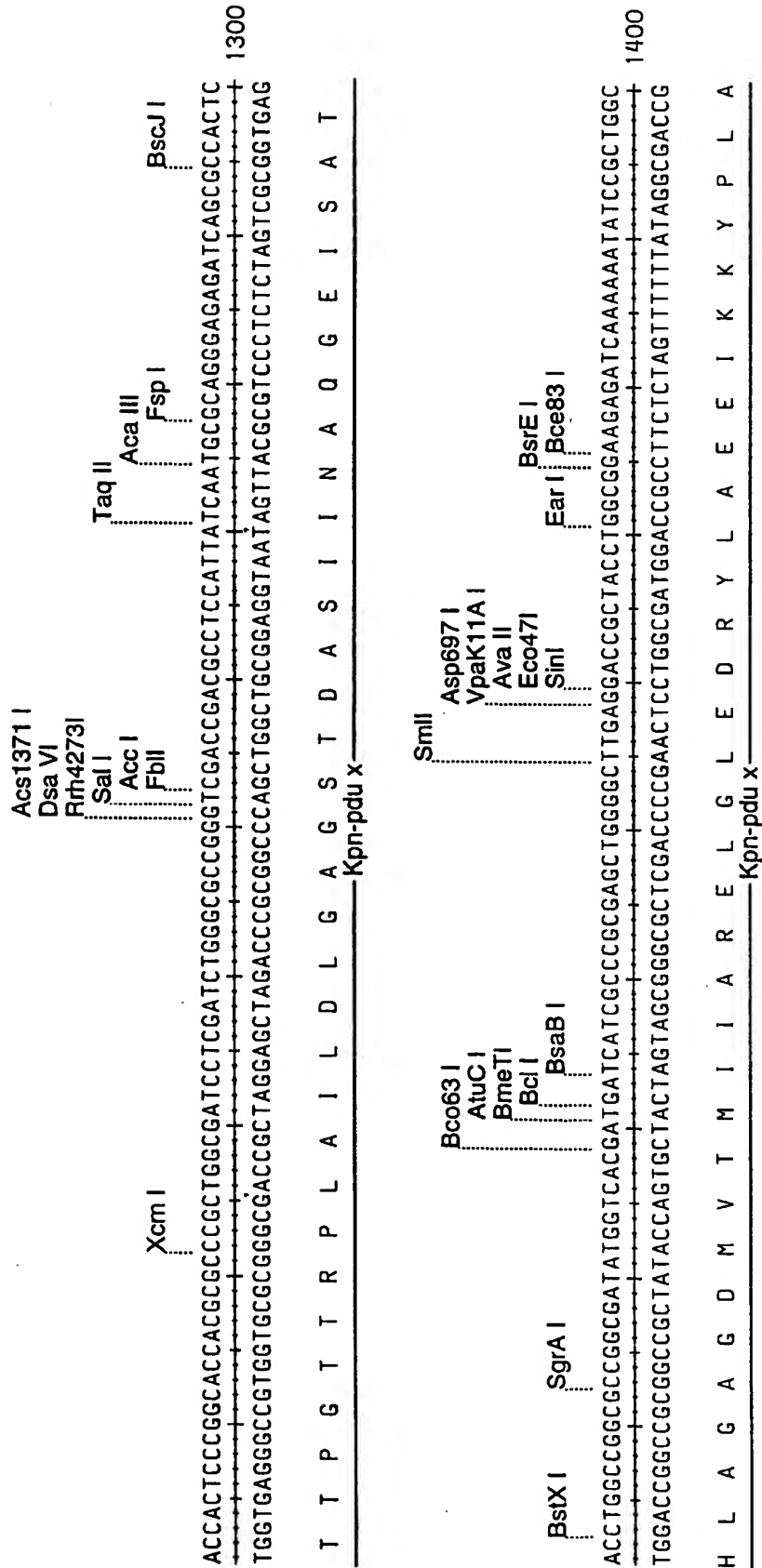


FIG.\_8D-1

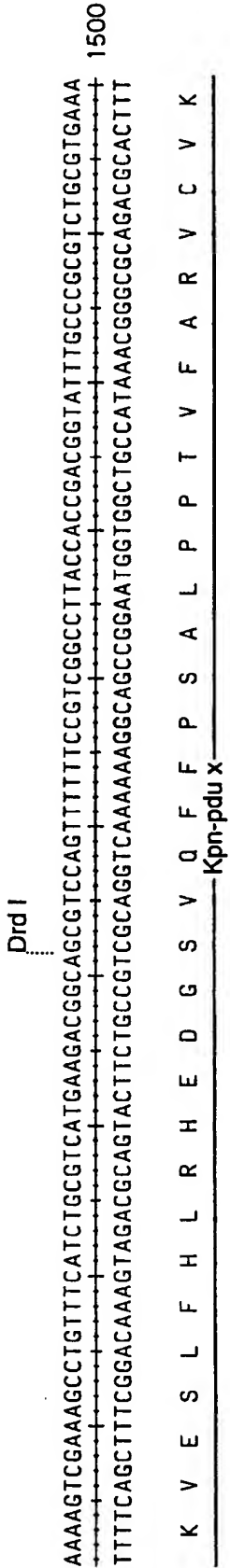


FIG.\_8D-2

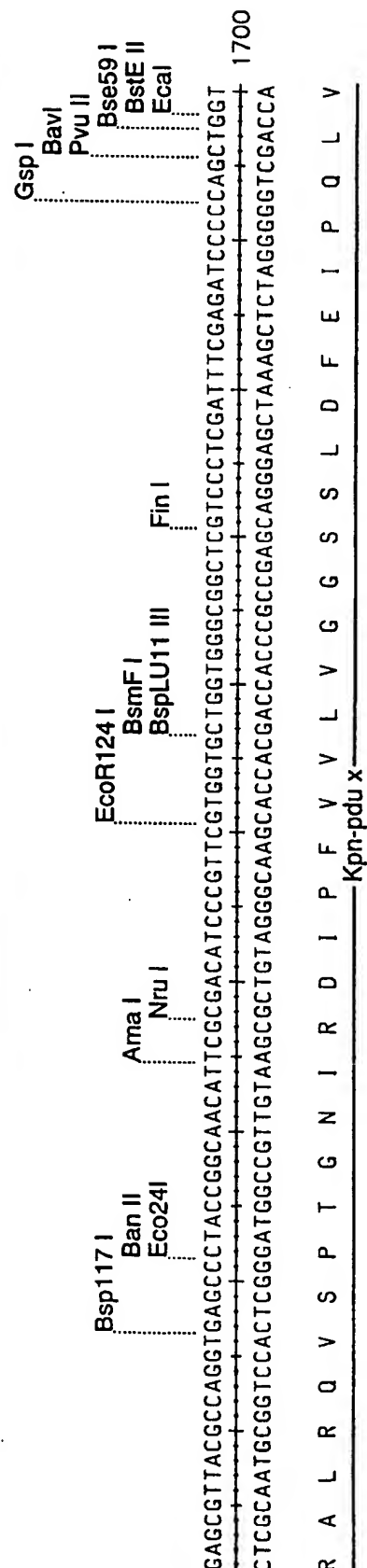
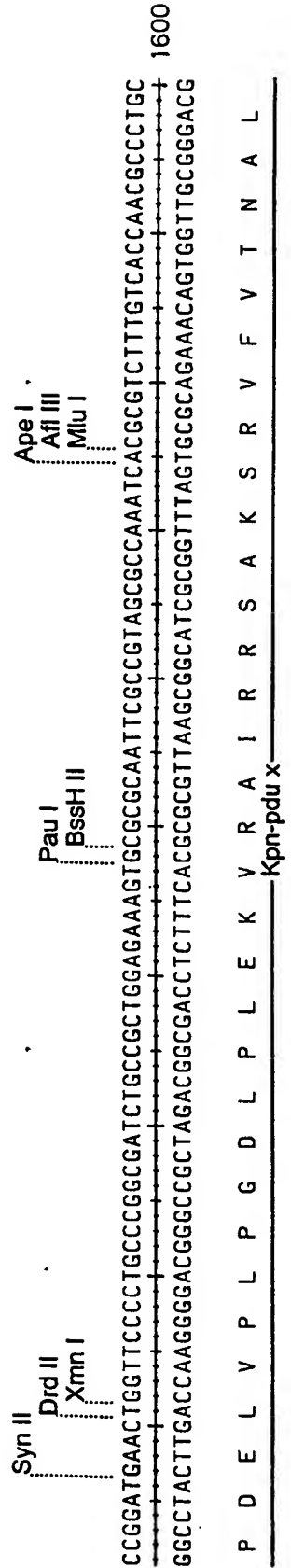


FIG.\_8E-1

26 / 27

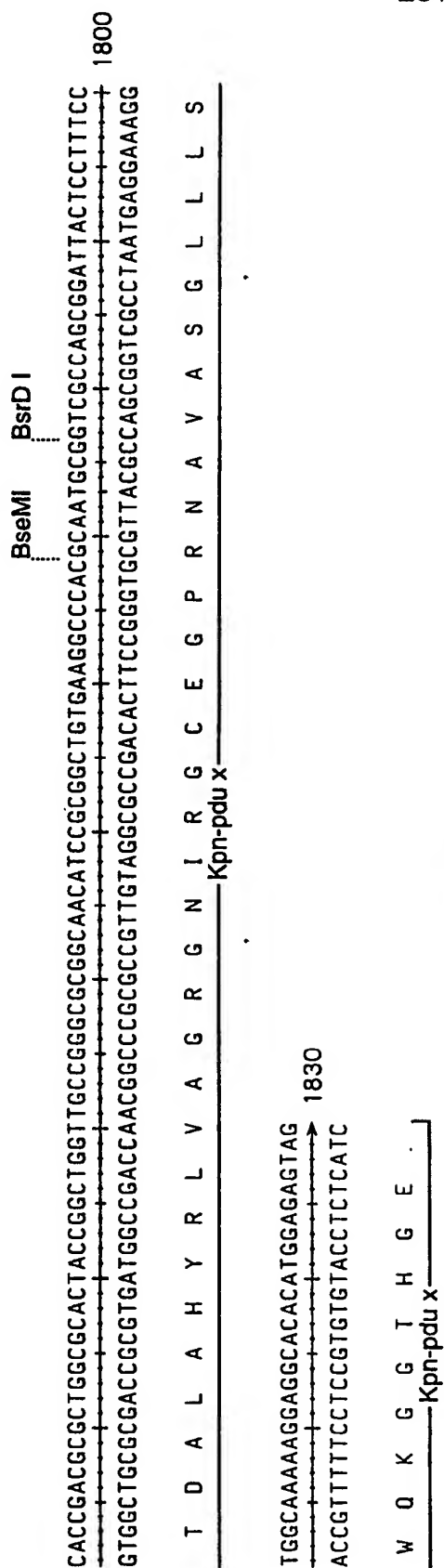
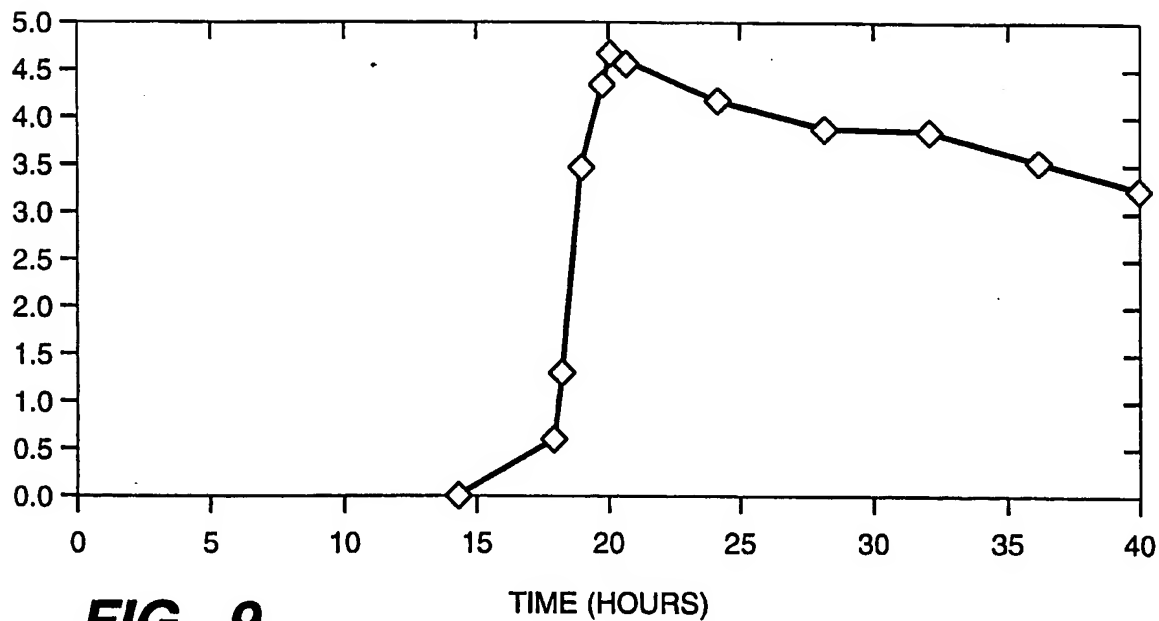
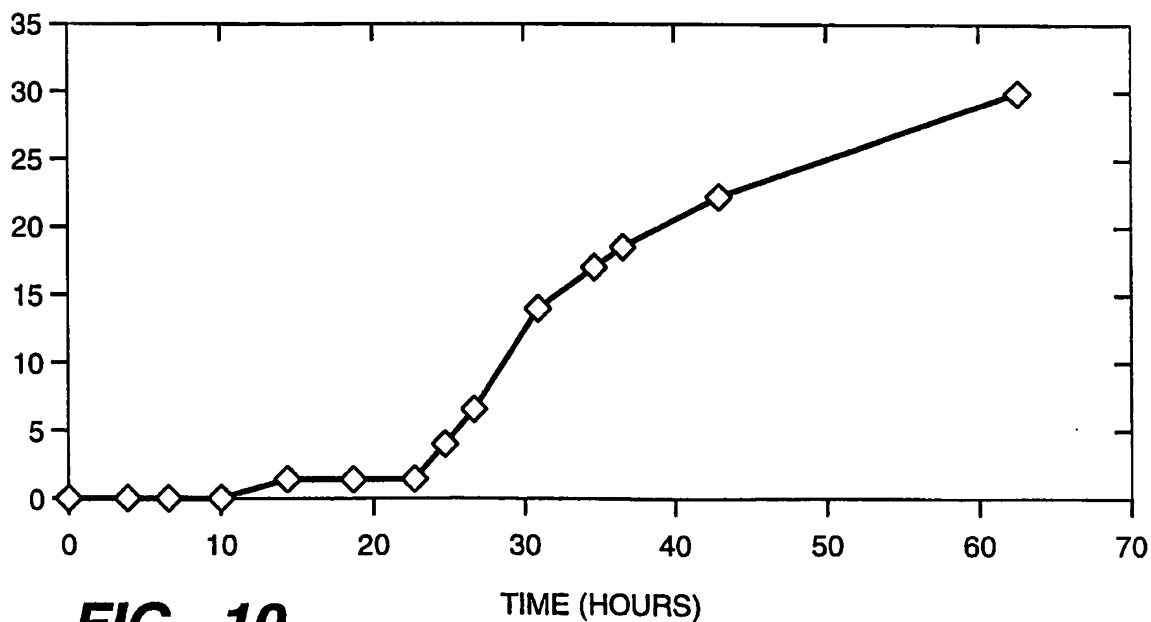


FIG. 8E-2



27 / 27

**FIG.\_9****FIG.\_10**

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>C12N 15/53, 15/55, 15/60, C12P 7/18,</b> <b>C12N 9/04, 9/16, 9/88</b>	<b>A3</b>	<b>(11) International Publication Number:</b> <b>WO 98/21341</b>
		<b>(43) International Publication Date:</b> 22 May 1998 (22.05.98)
<b>(21) International Application Number:</b> PCT/US97/20873 <b>(22) International Filing Date:</b> 13 November 1997 (13.11.97)  <b>(30) Priority Data:</b> 60/030,601 13 November 1996 (13.11.96) US  <b>(71) Applicant (for all designated States except US):</b> GENENCOR INTERNATIONAL, INC. [US/US]; 4 Cambridge Place, 1870 South Winton Road, Rochester, NY 14618 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> DUNN-COLEMAN, Nigel, S. [GB/US]; 142 Johnson Avenue, Los Gatos, CA 95032 (US). DIAZ-TORRES, Maria [ES/US]; 58 North El Camino Real, San Mateo, CA 94401 (US). CHASE, Matthew, W. [US/US]; 2211-27 Hastings Drive, Belmont, CA 94002 (US). TRIMBUR, Donald [US/US]; 349 Orchard Avenue, Redwood City, CA 94601 (US).  <b>(74) Agent:</b> GLAISTER, Debra, J.; Genencor International, Inc., Page Mill Road, Palo Alto, CA 94304-1013 (US).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>  <b>(88) Date of publication of the international search report:</b> 25 June 1998 (25.06.98)
<b>(54) Title:</b> METHOD FOR THE RECOMBINANT PRODUCTION OF 1,3-PROPANEDIOL  <b>(57) Abstract</b>  The present invention provides an improved method for the production of 1,3-propanediol from a variety of carbon sources is an organism comprising DNA encoding protein X of a dehydratase or protein X in combination with at least one of protein 1, protein 2 and protein 3. The protein X may be isolated from a diol dehydratase or a glycerol dehydratase. The present invention also provides host cells comprising protein X that are capable of increased production of 1,3-propanediol.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/US 97/20873

<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 6    C12N15/53    C12N15/55    C12N15/60    C12P7/18    C12N9/04 C12N9/16    C12N9/88		
According to International Patent Classification (IPC) or to both national classification and IPC		
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 6    C12P		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the international search (name of data base and, where practical, search terms used)		
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X, L	WO 96 35796 A (DU PONT ; GENENCOR INT (US); LAFFEND LISA ANNE (US); NAGARAJAN VASA) 14 November 1996 see the whole document see abstract see examples 2-5 see examples 22, 23 see page 62, paragraph 2 ---	1-40
P, A	WO 96 35795 A (DU PONT ; NAGARAJAN VASANTHA (US); NAKAMURA CHARLES EDWIN (US)) 14 November 1996 see abstract see page 9 see examples 1-3 ---	1-13, 15-40
-/--		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C. <input checked="" type="checkbox"/> Patent family members are listed in annex.		
* Special categories of cited documents : <div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <p>"A" document defining the general state of the art which is not considered to be of particular relevance</p> <p>"E" earlier document but published on or after the international filing date</p> <p>"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</p> <p>"O" document referring to an oral disclosure, use, exhibition or other means</p> <p>"P" document published prior to the international filing date but later than the priority date claimed</p> </div> <div style="width: 45%;"> <p>"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</p> <p>"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</p> <p>"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.</p> <p>"&amp;" document member of the same patent family</p> </div> </div>		
Date of the actual completion of the international search  <div style="text-align: center; font-size: 1.2em;">21 April 1998</div>		Date of mailing of the international search report  <div style="text-align: center; font-size: 1.2em;">08/05/1998</div>
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651-epo nl, Fax: (+31-70) 340-3016		Authorized officer  <div style="text-align: center; font-size: 1.2em;">Lejeune, R</div>

## INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 97/20873

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	<p>TOBIMATSU T. ET AL.: "Molecular cloning, sequencing, and expression of the genes encoding adenosylcobalamin-dependent diol dehydrase of <i>Klebsiella oxytoca</i>." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 270, no. 13, 31 March 1995, pages 7142-7148, XP002062740 cited in the application see page 7147, column 2, paragraph 2 see page 7145, column 1, line 1 - page 7146, column 1, line 4 see abstract</p> <p>---</p>	1,5,7,8, 10,11
A	<p>TOBIMATSU T. ET AL.: "Cloning, sequencing and high level expression of the genes encoding adenosylcobalamin-dependent glycerol dehydrase of <i>Klebsiella pneumoniae</i>." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 271, no. 37, 13 September 1996, pages 22352-22357, XP002057923 see abstract see page 22356, column 1, line 2 - line 5 see page 22356, column 2, paragraph 4 - page 22357, column 1, paragraph 1</p> <p>---</p>	1,4,6,8, 10,11,15
A	<p>TONG I.T. ET AL.: "1,3-propanediol production by <i>Escherichia coli</i> expressing genes from the <i>Klebsiella pneumoniae</i> dha regulon." APPLIED AND ENVIRONMENTAL MICROBIOLOGY, vol. 57, 1991, pages 3541-3546, XP002062741 see abstract</p> <p>---</p>	1-4,6,8, 10,11
A	<p>SEYFRIED M. ET AL.: "Cloning, sequencing, and overexpression of the genes encoding coenzyme B12-dependent glycerol dehydratase of <i>Citrobacter freundii</i>." JOURNAL OF BACTERIOLOGY, vol. 178, no. 19, October 1996, pages 5793-5796, XP002062742 cited in the application see page 5794; figure 1 see page 5793, column 2, paragraph 5 see abstract</p> <p>-----</p>	1-4,6

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 97/20873

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9635796 A	14-11-96	US 5686276 A AU 5678996 A EP 0826057 A	11-11-97 29-11-96 04-03-98
WO 9635795 A	14-11-96	US 5633362 A AU 5722996 A EP 0827543 A	27-05-97 29-11-96 11-03-98